Genome Organization in Sponges

Thesis

UNRESTRICTED

Genome Organization in Sponges

Maria Costantini

Laurea in Scienze Biologiche

Università degli Studi di Napoli "Federico II", Italia

Doctor of Philosophy

Sponsoring Establishment
Stazione Zoologica "Anton Dohrn"
Napoli, Italia

September 2003

Submission date: 15 September 2003
Award date: 30 December 2003

Director of Studies: Prof. Giorgio Bernardi

Superadvisor: Prof. Roberto Di Lauro

External Superadvisor: Prof. Charles Auffray

External Examiner: Prof. Michael Thorndyke

Internal Examiner: Prof. Mario De Felice

- Acknowledgements -

My foremost thanks go to Prof. Giorgio Bernardi, without whom this thesis would not have been possible. Throughout my thesis-writing period, he provided encouragement, sound advices, helpful teaching and stimulating ideas. I thank Giorgio Matassi for his tutoring of my thesis. I am grateful to Giuseppe D'Onofrio who convinced me, at the beginning of my thesis work, of the power of the CsCl profile parameters for phylogeny.

I am indebted to many people of my laboratory who taught me to use the tools I have applied in this thesis. First of all, I would like to thank Giuseppe Bucciarelli and Giuliana de Luca di Roseto for introducing me to the practical techniques in our laboratory. I would like to thank Oliver Clay and Nicolas Carels for very frequent discussions on the interpretation of the CsCl analytical profiles and for clarifying many technical questions regarding isochores and isochore families. I gratefully acknowledge Stilianos Arhondakis, Fabio Auletta and Giuseppe Torelli for their help in retrieving compositional content of coding sequences.

I would like to take this opportunity to extend my thanks also to Hector Musto, Fernando Alvarez-Valin, Kamel Jabbari and Adam Pavlíček for several discussions on diverse topics relevant to this thesis.

I want to thank Miriam Di Filippo for her moral support during my experimental work. Thanks are also due to our technician, Salvatore Bocchetti, for his support in the analytical ultracentrifugation.

Many people at the Stazione Zoologica gave other important contributions to my thesis: Elisabetta Tosti, Marina Montresor and their group for cellular analysis, Alfonso Toscano's group and the fishing service of our Institute, in particular Pasquale Sansone, for the careful maintenance of the sponges.

Finally, I want to thank Gianna Di Gennaro and Romy Sole, for their extreme availability and for being so close to me in all circumstances during these four years.

Index

Chapter 1

- Introduction -

1.1 Human genome organization

The most elementary property of the genome is the nucleotide composition of the DNA. Its variation along the chromosome (heterogeneity) has been used over the years in our laboratory to study the organization of the genome in a number of eukaryotes. Moreover, the heterogeneity of base composition is also an extremely useful parameter for evolutionary studies (see below).

From CsCl density gradient ultracentrifugation analysis of genomic DNA, used to study GC variation, several informations can be extracted for genomics and evolutionary studies. GC is defined as the molar fraction of guanine and cytosine in a molecule or segment of DNA (the proportion of its base pairs that are GC rather than AT). This most fundamental base compositional property of double-stranded DNA can be easily measured in an analytical ultracentrifuge (Clay et al., 2003a). The measurements are made in density gradients of heavy salts. Of these salts, cesium chloride is the most widely used. It is commercially available in optical-grade quality, it allows a faithful (linear) portrayal of GC distributions in an analytical centrifuge (AUC), and it permits high-resolution fractionation according to GC content in a preparative ultracentrifuge. The technique of density gradient ultracentrifugation was introduced in 1957 by Meselson, Stahl and Vinograd. The principle is simple: a heavy salt of low molecular weight in solution will, upon centrifugation,

establish a density gradient. At sedimentation equilibrium, double-stranded DNA molecules having a given GC will be found neither at the meniscus nor in the pellet, but in a narrow band within the density gradient. One therefore places the DNA together with the salt solution in the ultracentrifuge cell, and allows salt and DNA to reach equilibrium, which under standard conditions is attained within 24 hours. The GC level of the DNA can be read from its position in the cell. Soon after the first experiments, it was discovered (Sueoka et al., 1959; Marmur and Doty, 1959; Rolfe and Meselson, 1959: Schildkraut et al., 1962) that, in CsCl gradients, the GC level of a double-stranded DNA molecule exhibits a remarkably linear relationship to the position of the molecule at sedimentation equilibrium. More precisely, the GC level of the DNA molecule is linearly related to the density of the CsCl solution at its equilibrium position. This density is called buoyant density and is measured from the radial distance from the ultracentrifuge axis. One can therefore measure not only the GC level of a sample of compositionally similar molecules, but also the GC distribution of compositionally similar molecules, which spans in the human genome a GC range from just under 30% to just over 60% GC (at scales up to several megabases). Indeed the CsCl absorbance profile of high molecular weight DNA fragment is, after a linear transformation of the horizontal axis, to a very good approximation, the GC distribution of the fragment. Only when the fragment is smaller than about 15 kb (10×10^6 Daltons) does diffusion seriously distort the profile. Similarly only when DNA fragments are heavily methylated or otherwise modified (as in T-seven phages), highly repetitive, or denatured do they shift from their expected equilibrium positions.

The power of the density gradient ultracentrifugation methodology is precisely that it allows DNA sequence information to be logically inferred without seeing the DNA sequence. In fact, the CsCl method has been of central importance in understanding compositional variation along mammalian chromosomes; some of the main conclusions were drawn well before any DNA sequences were known (Filipski et al., 1973; Thiery et al., 1976; Macaya et al., 1976). An early result was the discovery that mammalian genomes are organized into long, compositionally fairly homogeneous regions, called isochores. By comparing absorbance profiles of the same species for different fragment sizes (molecular weights), and by monitoring the profiles' resistance to narrowing as the fragment sizes are decreased, one can infer statistical properties of the mosaic GC variation along its chromosomes (Macaya et al., 1976; Cuny et al., 1981; Clay e al., 2001).

In the case of the human genome, the Gaussian components of the CsCl profile were called the "major components" and relative amounts of DNA were called the "compositional pattern" of the genome. In the human DNA profile (Fig. 1.1) four components can be identified L, H1, H2, H3, which represent 62.9%, 24.3%%, 7.5%, 4.7% of the genome, respectively. The remaining DNA corresponds to satellite and ribosomal sequences (Bernardi et al., 1985; Zerial et al., 1986; Zoubak et al., 1996). These components are made up of large DNA segments, more than 300 kb in size, called isochores (Cuny et al., 1981) and arranged in a mosaic-like fashion along the chromosome. Isochores are compositionally homogeneous regions. Compositional homogeneity of isochores means that the GC heterogeneity within an isochore is much smaller than the heterogeneity among isochores.

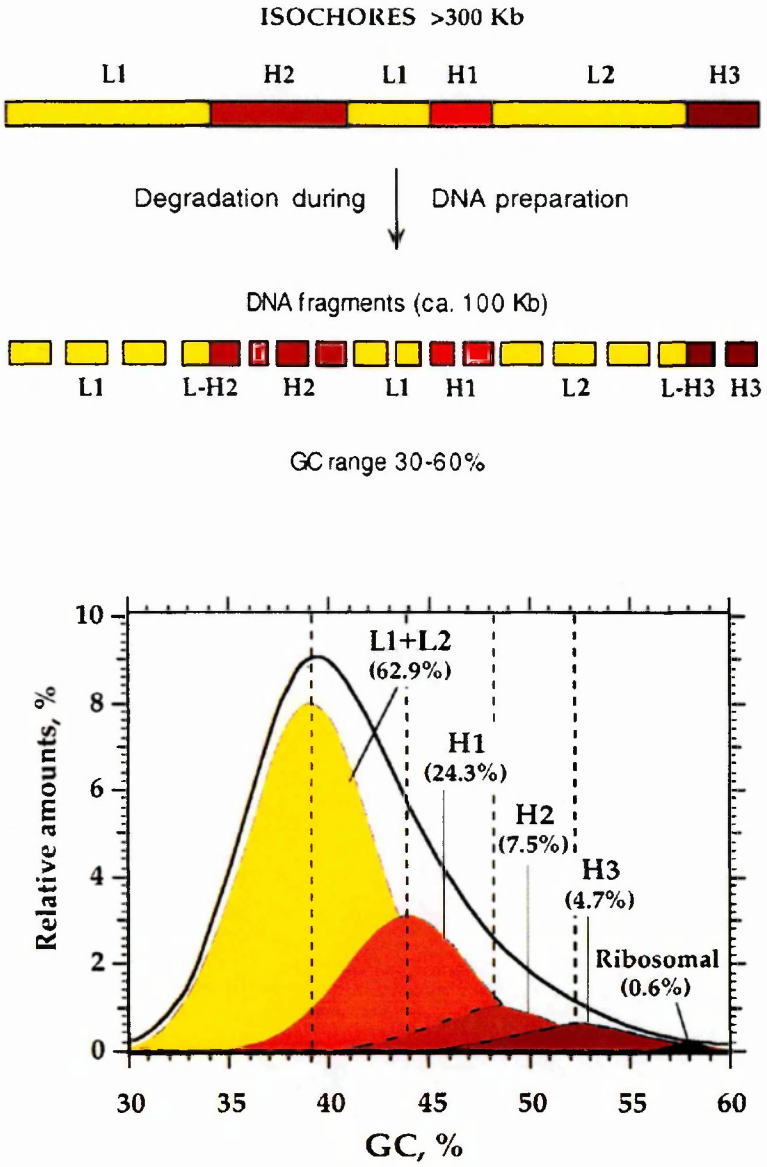


Fig. 1.1 (Top) Scheme of the isochore organization of the human genome. This genome, which is typical of the genome of most mammals, is a mosaic of large (>>300 kb, on average) DNA segments, the isochores, which are compositionally homogeneous (above a size of 3kb) and can be partitioned into a number of families. Isochores are degraded during routine DNA preparations to fragments of approx. 100 kb in size. The GC-range f the isochors from the human genome is 30-60% (from Bernardi 1995). (Bottom) The CsCl profile of human DNA is resolved into its major DNA components, namely the families of DNA fragments derived from isochore families L (i.e., L1+L2), H1, H2, H3. Modal GC levels of isochore families are indicated on the abscissa (broken vertical lines). The relative amounts of major DNA components are indicated. Satellite DNAs are not represented (from Zoubak et al., 1996).

The heterogeneity of the base composition is a crucial parameter to study the organization of the eukaryotic genome and for evolutionary analyses. For example it is important to distinguish between the highly heterogeneous genomes of warm-blooded vertebrates and the much less heterogeneous genomes of cold-blooded vertebrates: Fig. 1.2 shows that the isochore patterns are remarkably different in cold- and warm-blooded vertebrates.

Isochores, i.e. genome compartments, have both structural and functional significance. An obvious question is whether there is any correlation between the compositional patterns of coding sequences (which represent as little as 3% of the genome in vertebrates) and the compositional patterns of DNA fragments (97% of which are formed by intergenic sequences and introns). Another question is whether there is any correlation within genes between the composition of the exons and that of introns.

Indeed, linear correlations hold between the GC levels (and the GC_3 levels) of coding sequences and the GC levels of isochores in which coding sequences are located (see Fig. 1.3a, c). Interestingly, GC-poor coding sequences and their flanking sequences show very similar values, whereas GC-rich coding sequences are increasingly higher above the diagonal, essentially because GC_3 values depart more and more from the intergenic sequences (Fig. 1.3c). Linear correlations (Fig. 1.3) also hold betweeen the GC levels of coding sequences and the GC levels of the introns of the same genes (Bernardi et al., 1985; Aïssani et al., 1991; Clay et al., 1996), the GC levels of the former being slightly higher than those of the latter. These differences are much larger in plants (Carels and Bernardi, 2000). As a final remark, one should note that the correlations of Fig. 1.3a and b are

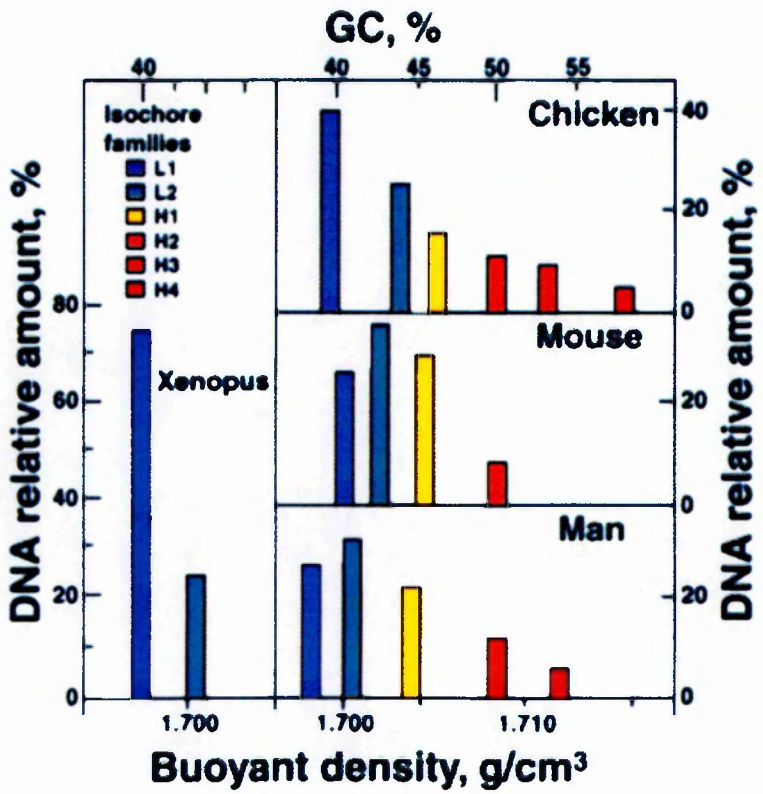


Fig. 1.2 Compositional patterns of vertebrate genomes. Histograms showing the DNA relative amounts, modal buoyant densities and modal GC levels of the major DNA components (the families of DNA fragments derived from different isochore families; see Fig. 1.1) from *Xenopus*, chicken, mouse and man, as estimated after fractionation of DNA by preparative density gradient. Satellite and minor DNA components (such as ribosomal DNA) are not shown. (Modified from Bernardi, 1995).

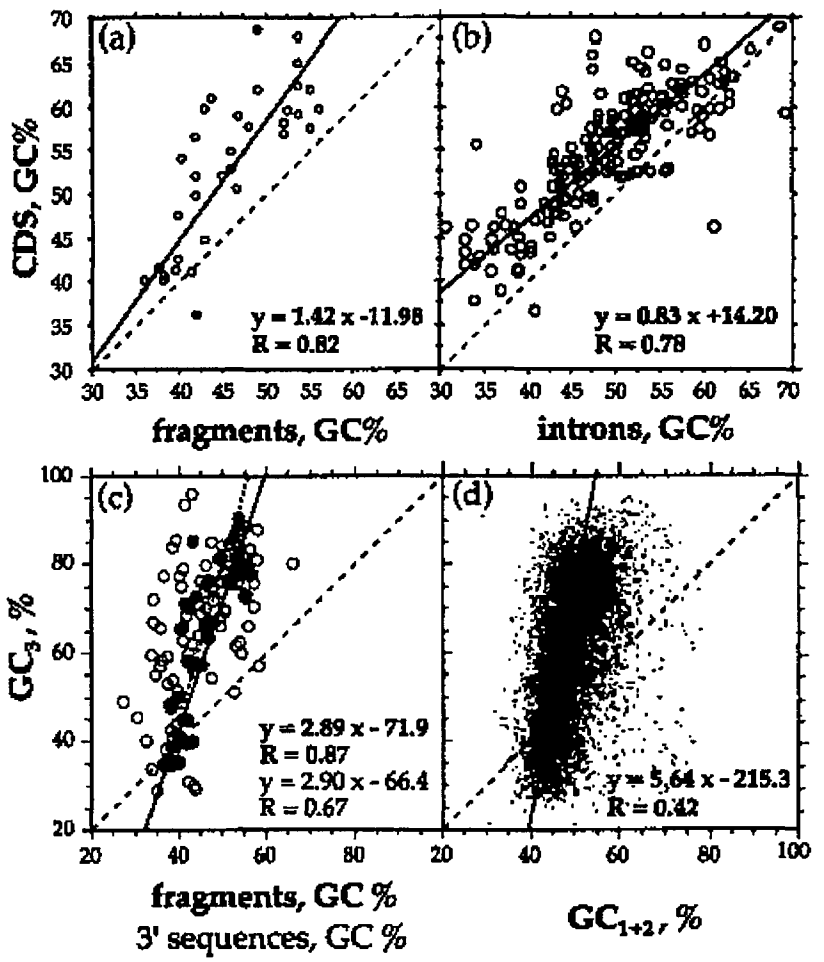


Fig. 1.3 Correlation between GC levels of human coding sequences and (a) the GC levels of the large DNA fragments in which sequences were localized, or (b) the GC levels of the corresponding introns (top frames). The bottom frames show the correlations between GC_3 of human coding sequences and (c) the GC levels of the DNA fractions in which the genes were localized (filled circles) and of 3' flanking sequences further than 500 bp from the stop codon (open circles; the solid and the broken lines are the regression lines through the two sets of points); or (d) GC_1 + GC_2 values of human sequences. Diagonals (unity slope lines) are also shown (from Clay et al., 1996).

practically the same in the chicken genome (Musto et al., 1999), and possibly in other vertebrate genomes.

The correlation between GC_3 levels of coding sequences and GC levels of isochores (Fig. 1.3c) is especially important, because it allows the positioning of the distribution profile of coding sequences relative to that of DNA fragments, the CsCl profile. In turn, this allowed us to estimate the relative gene density by dividing the percentage of genes located in given GC intervals by the percentage of DNA located in the same interval. Since it had been tacitly assumed that genes were uniformly distributed in eukaryotic genomes, it came as a big surprise that the gene distribution in the human genome is strikingly non-uniform (Fig. 1.4), gene concentration increasing from a very low average level in L isochores to a 20-fold higher level in H3 isochores (Bernardi et al., 1985; Mouchiroud et al., 1991; Zoubak et al., 1996). The existence of a break in the slope of gene concentration at 60% GC_3 of coding sequences and at 46% GC of isochores (see Fig. 1.4) defines two "gene spaces" in the human genome. In the "genome core" (Bernardi, 1993a, 1995), formed by isochore families H2 and H3 (which make up 12% of the genome), gene concentration is very high (one gene per 5-15 kb) and comparable to those of compact genomes of higher eukaryotes, whereas in the "empty space", formed by isochores families L and H1 (which make up 88% of the gnome) gene concentration is very low (one gene per 50-150 kb). Fig. 1.5 represents the density of gene sequences in isochore families. About 54% of human genes are located in the small "genome core", the remaining 46% being located in the large "empty quarter".

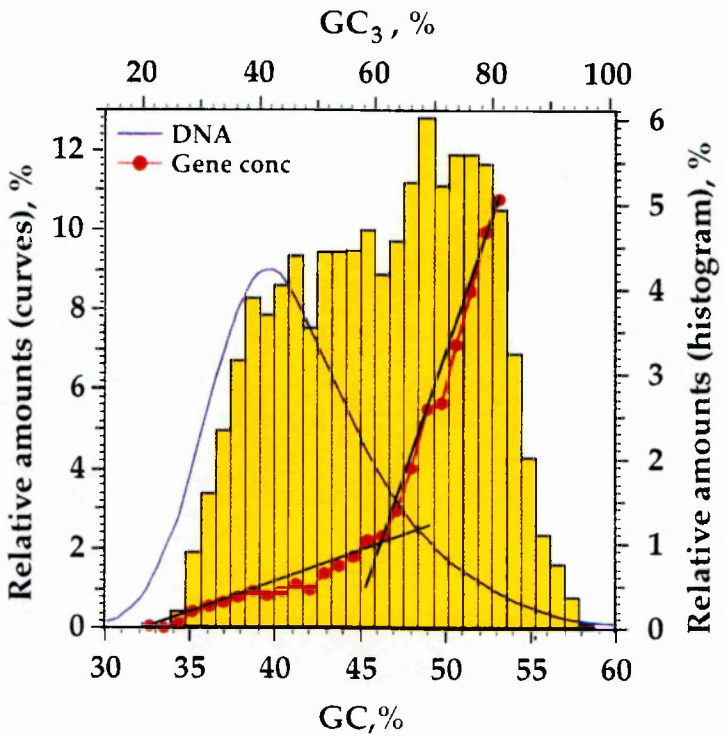


Fig. 1.4 Profile of gene concentration (red dots) in the human genome, as obtained by dividing the relative numbers of genes in each 2% GC_3 interval of the histogram of gene distribution (yellow bars) by the corresponding relative amounts of DNA deduced from the CsCl profile (blue line). The positioning of the GC_3 histogram relative to the CsCl profile is based on the correlation of Fig. 1.3c. The apparent decrease in the concentration of protein-encoding genes for very high values (broken line) is due to the presence of ribosomal DNA in that region. The last concentration values are uncertain because they correspond to very low amounts of DNA (from Zoubak et al., 1996).

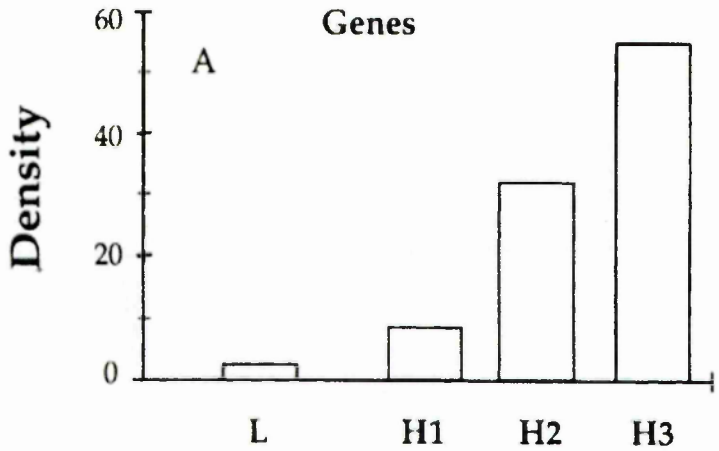


Fig. 1.5 Density of gene sequences in isochore families. Relative numbers of sequences over relative amounts of isochore families are presented in the histograms (from Zoubak et al., 1996).

1.2 Sponges (Porifera)

The transition from unicellular to multicellular organisms occurred in all five kingdoms of life: this process took place impressively in Fungi (Ascomycota), Plantae (Chlorophyta) and in Metazoa (Müller, 1998). The origin of plants appears to be well elucidated within the phylum Chlorophyta (Margulis and Schwartz, 1995), while the origins of Fungi and especially of Metazoa are perhaps still the most enigmatic of all phylogenetic problems.

The evolution of Metazoa from their protozoan ancestors has been considered, until recently, as the greatest puzzle of phylogeny (Willmer, 1994; Cavalier-Smith, 1991). The emergence of metazoan has been explained by two major theories: the syncytial theory (origin from a multinucleated ciliate) (Hadzi, 1963), or the colonial theory (origin from a colonial flagellate) (Haeckel, 1868). However, a di(poly)phyletic origin of Metazoa is assumed in both cases.

The phylogenetic relationship of the kingdom Animalia (Metazoa) has long been questioned. Initially, detailed descriptions of animal embryology and adult morphology were used to solve the evolutionary origins of distant groups such as phyla. Focusing on the lowest eukaryotic multicellular organisms, the metazoan phylum Porifera (sponges), it remained unclear if they independently evolved multicellularity from a separate protist lineage (polyphyly of animals) or derived from the same protist group as the other animal phyla (monophyly) (Müller, 1998). Based on constituent characters of the sponges a monophyletic origin of the Porifera can be deduced. The oldest complete fossil sponge has been described from the Early Cambrian, while the earliest spicules date from the late

Proterozoic, about 600 million years ago. It is suggested that the first sponges did not contain spicules. After having analyzed those genes from the sponge *Geodia cydonium*, which are typical for multicellularity, for example those coding for adhesion molecules/receptors and a nuclear receptor, it has to be concluded that all animals, including sponges, are of monophyletic origin. In this regards, *Geodia cydonium* might be considered as a "living fossil" not only suitable for the studies of adhesion molecules and receptors found in sponges and in eumetazoans, but also for the elucidation of other typical metazoan circuits for example functions in light-sensitive organs (βγ-crystallin has been cloned from *Geodia cydonium*) or the basis of the invertebrate immune system (immunoglobulin, subunits of proteasomes and heat shock proteins), as proposed by Müller (1997).

In fact, it should be stressed that evolution is a gradual process whereby new genes are formed primarily by either gene duplication (Ohno, 1970) or exon shuffling (Gilbert, 1978). In addition, new proteins can also be produced by overlapping genes, alternative splicing, or gene sharing (Li and Graur 1991). These facts imply that (a) proteins found for the first time in a given phylum contain elements, modules, which are present already in ancestral protein(s) of members of phylogenetically older phyla, and (b) that new combinations of modules create proteins that possess new functions.

Therefore Müller in 1998 postulated that animals, which are positioned at the base of Metazoa, such as sponges, are especially rich in ancestral modules for structural and functional molecules found also in higher Metazoa. This approach proved successful. As outlined, the structures of the characteristic metazoan genes and proteins required for (a)

tissue formation (galectin, collagen, integrin), (b) signal transduction (tyrosine kinase receptor RTK), (c) transcription (homeodomain and MADS box containing proteins), (d) immune reactions (heat shock proteins, proteasome, proteins featuring SRCR domains, and (e) sensory tissue (crystallin, glutammate receptor) have been identified in *Geodia cydonium* (Fig. 1.6) and found to display high similarity to sequences from members of higher metazoan phyla (Müller, 1997). Based on the available sequence data it is reasonable to place Porifera in the kingdom Animalia together with the Metazoa ((Müller et al., 1994; Müller, 1995; Müller, 1997). It addition, as taken from the first sponge genes, especially that coding for RTK, it is now established that modular proteins, formed by exon-shuffling, are common to all metazoan phyla. This mechanism of exon-shuffling is apparently absent in plants and protists (Patty, 1995). If this view can be accepted, the "burst of evolutionary creativity" during the period of the Cambrian explosion which resulted in the "big bang" of metazoan radiation (Lipps and Signor, 1992) was driven by the process of modularization. During this process the already existing domains were transformed into mobile modules allowing the composition of mosaic proteins (see Fig. 1.6).

In addition it was estimated that the adhesion molecules/receptors from sponges diverged from a common ancestor in the Precambrian, about 800 million years ago.

It was hoped that nucleotide sequence data from rRNA would help to solve the question of metazoan phylogeny. Applying this approach and excluding the lowest metazoan phylum, the Porifera (sponges), several authors have assumed that multicellular animals have evolved only once (Field et al., 1988; Lake, 1990).

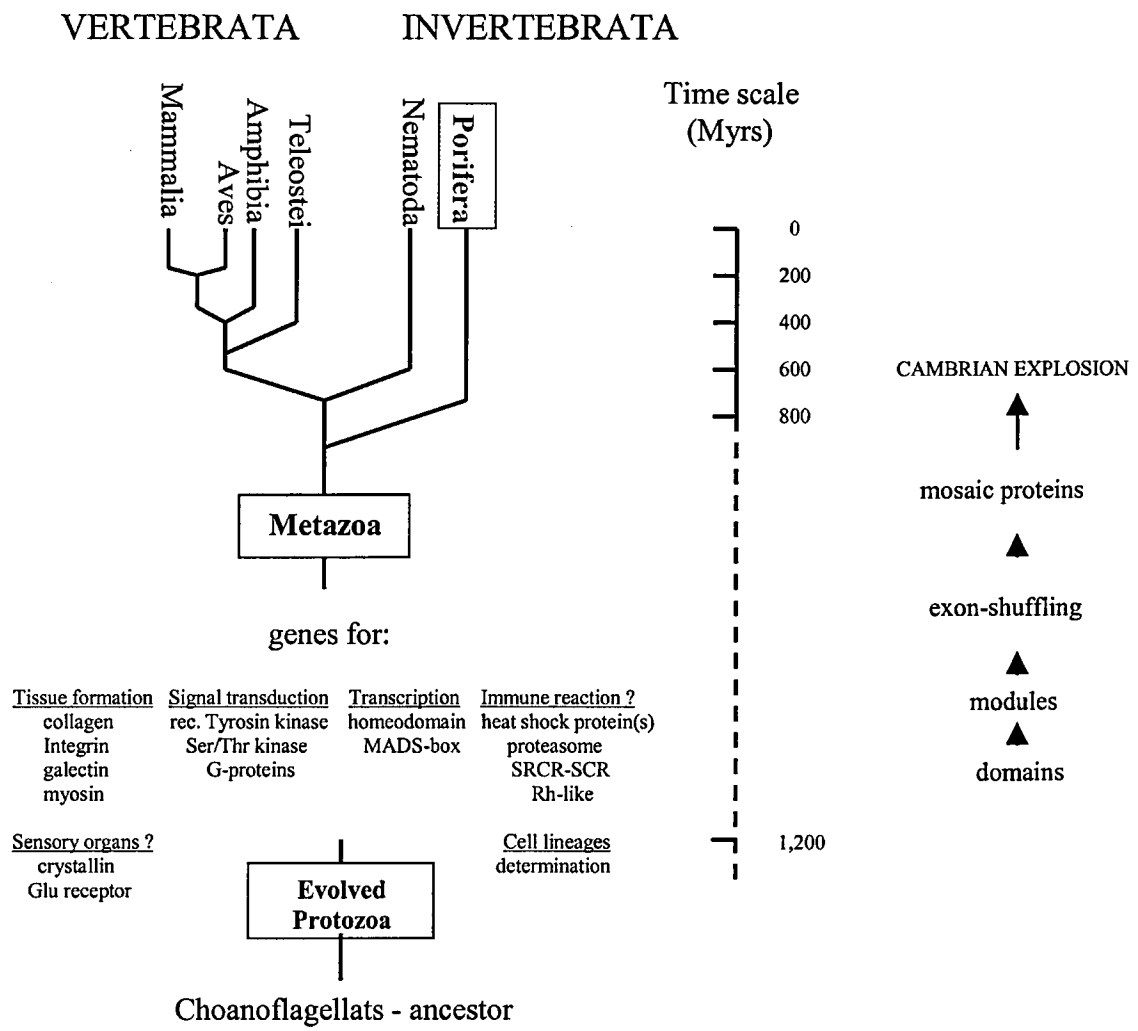


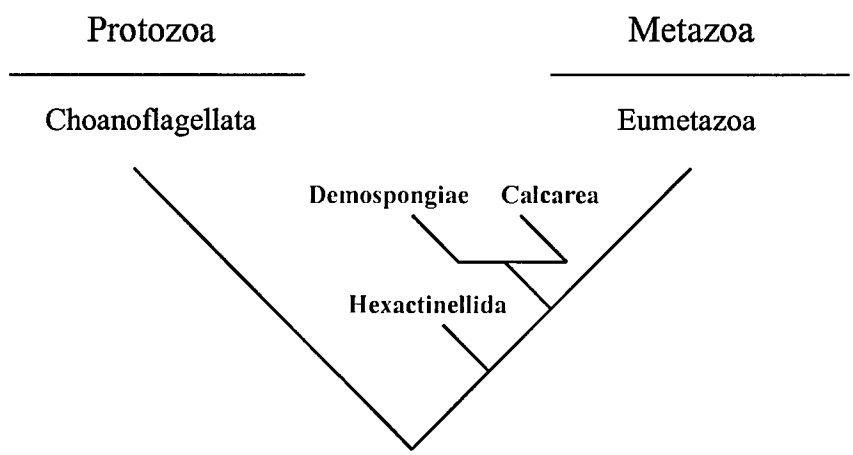
Fig. 1.6 Phylogenetic relationship of Porifera within the animal groups based on molecular biological data, obtained from sequences of "metazoan" proteins required for tissue formation, signal transduction, transcription, immune reaction (potential) and sensation (potential). It is proposed that the Cambrian explosion of metazoan radiation became possible after the creation of the evolutionary mechanism of modularisation of distinct protein domains, thus allowing the formation of mosaic proteins by exon-shuffling; this process happened approximately 1000million years ago. It is thought that Metazoa originated from evolved Protozoa, for example, Choanoflagellata. (Modified from Müller, 1998).

However, when sequences derived from 18S (Field et al., 1998) or 28S (Christen et al., 1991) rRNA from sponges are included, the assumption has been derived that the Radiata (including Porifera, Placozoa, Cnidaria and Ctenophora) and the Bilateria (other animal phyla) originated separately from different protozoan ancestors. Analyses of the 18S rRNA sequence have proved unsuitable for resolving deep branching in the phylogenetic tree, such as the positioning of the phylum Porifera within the kingdom of Metazoa (Rodrigo et al., 1994).

Willmer (1994) has pointed out that only a few (perhaps only two) developmental strategies would have allowed the transition from Protists to Metazoa; first, by aggregation of either mitotically related or unrelated cells, and second, by the formation of multinucleate cells after incomplete division of the cytoplasm. In both cases, the metazoan ancestor must have acquired the ability of interactions (1) between cells and (2) subsequently also between cells and the extracellular matrix.

Two alternative hypotheses have been proposed to explain the relationships between the major sponge classes. There are three sponge classes: Hexactinellidae, Demospongiae and Calcarea. One groups the Porifera into the adelphotaxa Hexactinellidae and Demospongiae/Calcarea (Fig. 1.7a) based on the gross difference in tissue structure and on differences in the structure of the flagella, whose beating generates the feeding current through sponges (Mehl and Reiswig, 1991). The other hypothesis assumes that the Demospongiae are more closely related to Hexactinellidae (Fig. 1.7b) based on presumed larval similarities (Boger, 1988).

a)



b)

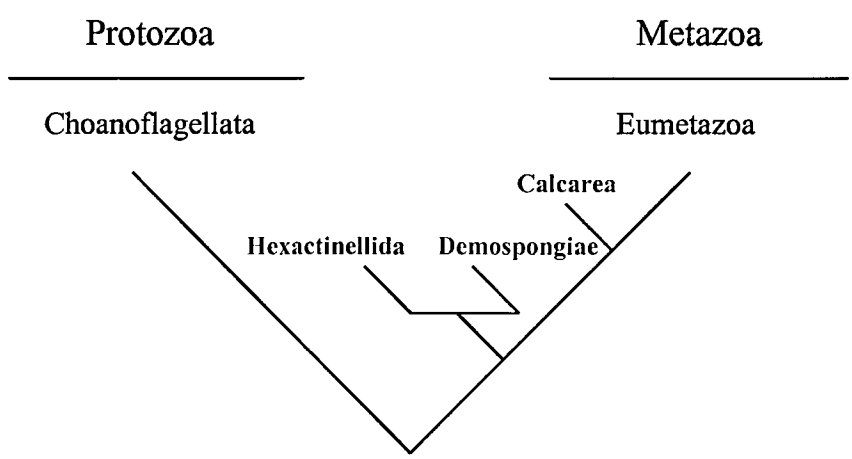


Fig. 1.7 Phylogenetic position between the major sponge classes: a) one hypothesis groups the Porifera into the adelphotaxa Hexactinellidae and Demospongiae/Calcarea, based on the gross difference in tissue structure and on differences in the structure of the flagella, whose beating generates the feeding current through sponges (Mehl and Reiswig, 1991); b) the other hypothesis assumes that the Demospongiae are more closely related to Hexactinellidae based on presumed larval similarities (Boger, 1988). (Modified from Müller, 1998).

The natural environmental factors exert strong pressure on the sponges. The success/failure to adapt to these various environmental conditions is one major factor that drives natural selection.

A critical parameter permitting the appearance of sponges was apparently oxygen. The emergence of metazoans and hence of Porifera as the first phylum, coincides with the increase in the atmospheric oxygen concentration from 10% to 100% of the present oxygen concentration in the atmosphere (Canfield and Teske, 1998). It may be proposed that the oxygenation of water is correlated with its use for collagen biosynthesis, for the hydroxylation of amino acids, one of the main novelties introduced by the sponges to the metazoan kingdom. The oxygen supply in sponges is maintained by the circulation of water through the efficient aquiferous channel system; it has recently been proposed that oxygen is a morphogenetic factor in these animals (Perovic et al., 2003). Besides oxygen, the supply of calcium ions (Ca^{2+}) is critical for metazoan animals. This ion is not only required for intracellular signal transduction but also for the establishment of cell-cell contacts, especially in sponges (Weinbaum and Burger, 1973; Müller and Zahn, 1973). The increase of Ca^{2+} in the oceans to the present-day level of $> 10^{-3}$ M only became possible after a decline in the alkalinity (Kemp and Kazmierczak, 1994).

Even though sponges inhabit almost all the substrata in the oceans from the Arctic to the Tropics (van Soest, 1994) to depths of over 2.000 m (Mehl, 1992), they can become very old (Gatti, 2002) and have been extremely successful survivors in Earth's history, they are sensitive to the effects of climate and anthropogenic changes. As a major factor, temperature increase can be postulated (Perez et al., 2000) as leading, for example, to mass

mortality events during the last few decades in the Mediterranean Sea (Pronzato, 1999). It is obvious, especially in tourist areas that the diversity of sponges has declined and continues to decline. Some sponges have the unique ability to etch the calcareous substratum and to penetrate into it. In particular, the species of the genus *Cliona* are well known for their ability to dissolve calcium carbonate and to excavate, burrow, or bore into calcitic/aragonitic substrata. The effective enzyme (carbonic anhydrase) was localized on the outer surface of the etching cell on the filopodia and between cell processes (Pomponi, 1979). It was hypothesized that the enzyme is secreted into the surrounding milieu (Rützler and Rieger, 1973).

Sponges are able to completely change their survival strategies, for example according to the food supply (carnivorous nutrition; Vacelet and Boury-Esnault, 1995) and to contribute to the stability of whole ecosystems, such as coral reefs, thus providing a major key to understanding the "coral reef paradox" (Richter et al., 2001).

The topic for an extensive number of studies has been the fact that the sponge fauna changes within an area strongly dependent on the surface of the ground where they attach (see Vatova, 1928; Rützler, 1965) and perhaps on the inorganic components in the surrounding water. This fact contributes to the overall species diversity of this taxon and perhaps also to the speed of the process of speciation, but also implies the inherent danger that well-adapted species may become extinct.

At one time, a diagnostic feature of the Porifera was the presence of spicules. The Hexactinellidae, or glass sponges, are characterized by siliceous spicules consisting of six rays intersecting at right angles. In particular, much of their tissues are syncitia, extensive

regions of multinucleate cytoplasm. Some discrete cell types do exist, including archaeocytes. Whereas other sponges possess the ability to contract, hexactinellidae do not. Hexactinellidae possess a unique system for rapidly conducting electrical impulses across their bodies, allowing them to react quickly to external stimuli. The Demospongiae are by far the most diverse sponge group. They are the most widespread and advanced class of sponges: greater than 90% of the 5,000 known living sponge species are demospongiae. However, the vast majority of living demospongiae do not possess skeletons that would easily fossilize, thus their fossil diversity, which peaks in the Creataceous, is probably an enormous underestimate of their true diversity. As their great number of species would suggest, demospongiae are found in many different environments, from warm high-energy intertidal settings to quiet cold abyssal depths. Indeed, all of the known freshwater poriferans are demospongiae. Demosponge skeletons are composed of spongin fibres and/or siliceous spicules, though one genus (Oscarella) has neither. Demosponge spicules, if present, are siliceous, have one to four rays not at right angles, and have axial canals that are triangular in cross section. Members of the group Calcarea are the only sponges that possess spicules composed of calcium carbonate. These spicules do not have hollow axial canals. Today, their diversity is greatest in the tropics, as is the case with most marine groups, they are predominantly found in shallow waters, though at least one species is known from a depth of 4,000 meters. The fossil record of the Calcarea indicates that it has always been more abundant in near-shore shallow water settings.

The Porifera are present both in the marine and the freshwater biotope. Some of them are able to filter their own body volume of water every 5s in order to extract edible material

(Vogel, 1977). The flow speed of the water in the inhalant and exhalant canals is high; an output velocity of 20 cm/s (Reiswig, 1971) has been estimated. They ingest particles of size between 5 and 50 μm through the cells of the mesohyl and the pinacoderm, and microparticles (0.3 to 1 μm) via the cells of the choanocyte chambers. A sponge specimen of 1 kg may filter about 24000 litres d^{-1} (Vogel, 1997). Nutrients are acquired by phagocytosis of bacteria that are removed from the water column. Considering this amazingly large amount of water and all the adverse factors contained in it, it is surprising that sponges have survived over 500 My (Müller, 2003). It is even more impressive that they could resist severe ice periods, for example during Proterozoic or Phanerozoic (Knoll and Carroll, 1999).

Sponges have a cellular grade of organization. They do not possess any structures that can be considered organs. Instead, sponge cells of various types are responsible for bodily functions, the day-to-day activities that sustain life. Many of most common types of cells are illustrated in the cartoon view of the wall of a sponge (Fig. 1.8). The pinacocytes are the "skin cells" of sponges. They line the exterior of the sponge body wall. They are thin, leathery and tightly packed together. Choanocytes are distinctive cells that line the interior body walls. These cells have a central flagellum that is surrounded by a collar of microvilli. It is their striking resemblance to the single-celled protists called choanoflagellates that make many scientist believe that choanoflagellates are the sister group to the Animals. Choanocytes are versatile cells. Their flagella beat to create the active pumping of water through the sponge, while the collars of the choanocytes are the primary areas that nutrients are absorbed into the sponge. Furthermore, in some sponges the choanocytes develop into

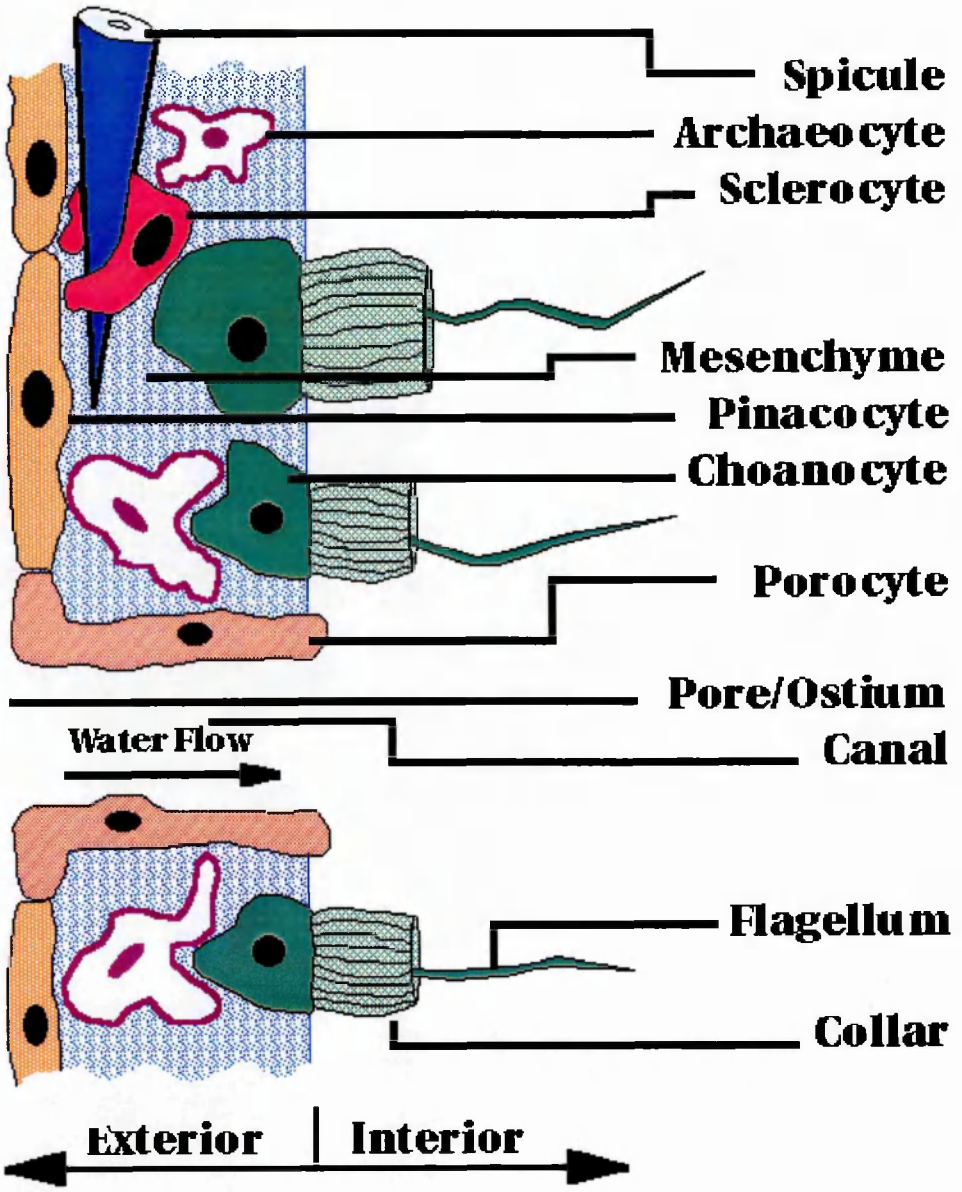


Fig. 1.8 Microscopic view of a poriferan wall. Many of the most common types of cells are illustrated in a cartoon view oh the wall of poriferan (available at www.ucmp.berkeley.edu/porifera/pororg.html).

gametes. Between the two layers is a thin space called mesenchyme or mesohyl. The mesenchyme consists of a proteinaceous matrix, some cells and spicules. Archaeocytes are very important to the functioning of a sponge. These cells are totipotent, which means that they can change into all of the other types of sponge cells. Archaeocytes ingest and digest food caught by the choanocyte collars and transport nutrients to the other cells of the sponge. In some sponges, archaeocytes develop into gametes. The secretion of spicules is carried out by sclerocytes. Other cells, called spongocytes, secrete the spongin skeletal fibres when those are present. Sponges do not have any muscle cells, so their movement is rather limited. However, some poriferan cells can contract in a similar fashion as muscle cells. Myocytes and porocytes which surround canal openings and pores can contract to regulate flow through the sponge.

The above characteristics of the sponge system make it attractive as a model for investigating basic mechanisms of cell-cell and cell-matrix interactions.

Reproduction by sponges is by both sexual and asexual means. Asexual reproduction is by means of external buds. Some species also reproduce from internal buds, called gemmules, which can survive extremely unfavourable conditions that cause the rest of the sponge to die. Sexual reproduction takes place in the mesohyl. Male gametes are released into the water by a sponge and taken into the pore system of its neighbours in the same way as food items. Spermatozoa are "captured" by collar cells, which then lose their collars and transform into specialized, amoeba-like cells that carry the spermatozoa to the eggs. Some sponges are monoecious; others are dioecious. In most sponges for which developmental patterns are known, the fertilized egg develops into a blastula, which is

released into the water. The larvae may settle directly and transform into adult sponges, or they may be planctonic for a time. Adult sponges are always sessile.

Sponges are known as rich sources of bioactive secondary metabolites. Sponges are thought to live in a symbiotic relationship with one-celled organisms such as prokaryotes, bacteria and primarily cyanobacteria (Vacelet, 1971) as well as eukaryotes, zooxantellae (yellow symbiotic dinomastigotes) (Sará and Liaci, 1964) or zoochlorellae (green symbiotic algae) (Gilbert and Allen, 1973). These organisms occur extracellularly and intracellularly (Wilkinson 1978). Antimicrobial compounds have been isolated from sponge-associated bacteria on numerous occasions, and this has prompted the suggestion that microbial symbionts play a role in the defence of their host sponge (Webster et al., 2001). Marine sponges produce a wide array of other natural products and bioactive secondary metabolites. The diversity of the secondary metabolites produced has been highlighted in a large number of reviews (Faulkner, 1995; Sarma, 1993). They range from derivatives of amino acids and nucleosides to macrolides, porphyrins, terpenoids to aliphatic cyclic peroxides and sterols. This diversity reflects the efficient mechanisms of combinatorial biochemistry which the animals have acquired during their evolutionary history. The question arises of whether the sponges, being the host of associated/symbiotic bacteria, are the producers or whether it is the microorganisms which they harbour (Müller et al., 2003). Recent data strongly favour the view that the microorganisms are the main producers of the natural products which are stored and accumulated in the sponge as a chemical mechanism (Proksch et al., 2002), although sponge metabolites can also be produced by specific sponge cells (Salomon et al., 2001): as an example, the phosphatase

inhibitor okadaic acid can be cited (Tachibana et al., 1981). This compound was first isolated from the sponge *Halicondria okadai* and was later found to be produced by the free-living microalgae *Prorocentrum lima* and perhaps even by bacteria which are associated with them (Murakami et al., 1982) Sponges such as *Suberites domuncula* use okadaic acid as defence against foreign eukaryotic organisms while at the same time they possess a relative resistance against this compound. Furthermore, *Suberites domuncula* takes advantage of the inhibitory activity of the compound by activating its MAP (mitogen-activated protein) kinase pathway (Wiens et al., 2003). For example *Vibrio* spp. associated with the sponge *Dysidea* sp. were shown to synthesize cytotoxic and antibacterial tetrabromodiphenyl ethers (Elyakov et al., 1991). The diketopiperazines associated with the sponge *Tedania ignis* were found to be produced by a *Micrococcus* sp. (Stierle et al., 1988). Recently, the antifungal peptide theopalauamide, isolated from the marine sponge *Theonella swinhoei*, was shown to be contained in a novel δ-proteobacterial symbiont (Schmidt et al., 2000). Some of these chemicals have been found to have beneficial pharmaceutical effects for humans, including compounds with respiratory, cardiovascular, gastrointestinal, anti-inflammatory, antitumor and antibiotic activities.

Despite their crucial position in evolution, there is not a lot of informations about the sponge genome. Using Feulgen staining the amount of DNA per cells has been estimated with 0.11 pg DNA in one sponge species, *Dysidea crawshagi* (Fasman, 1976). Applying the technique of flow cytometry and using DAPI as dye to stain the DNA quantitatively, the genome size of the haploid genome of marine sponges *Suberites domuncula* and *Geodia cydonium* results to be approximately 1.7 pg, corresponding to 1.7

x 10^9 bp. This value is in the range of those found in some vertebrates, for example *Gallus domesticus* (chicken) in which the genome size is 1.2 x 10^9 bp or *Cyprinus carpio* in which is 1.2 x 10^9 bp. In comparision, the size of the human haploid genome is 3.3 x 10^9 bp (Li and Graur, 1991). Chromosomes could only be visualized in the sponge *Suberites domuncula*. In the diploid state the karyotype of the *Suberites domuncula* is 32 chromosomes. They appear (Fig. 1.9) spherulous in shape under the microscope and their size is between 0.25 and 1.0 µm. (Imsiecke et al., 1995). In the prophase (Fig. 1.9a and b) the chromosomes are very thin (0.25 µm in maximum) and condense with time (0.5 µm). With transition to metaphase (Fig. 1.9c and d) the chromosomes reach their maximum density and thickness; they showed a spheric to rod-like shape (0.75 to 1.0 µm). In the early anaphase the chromosomes are obviously arranged into two groups of chromatids suggesting a spindle apparatus. In the late anaphase the chromosomes are separated into two different nuclei.

In comparison with chromosomes of the freshwater sponge *Spongilla lacustris* which have size between 0.7 and 2.1 µm (Imsiecke et al., 1993) the dimensions of the chromosomes from *Suberites domuncula* are smaller. It was not possible to identify unequivocally centromeres in the chromosome preparations from *Suberites domuncula*; the same difficulty was noticed already with the description of the chromosomes from *Spongilla lacustris*. A distinct banding pattern of the sponge chromosomes is not visible. No chromosomes could be identified in *Geodia cydonium*.

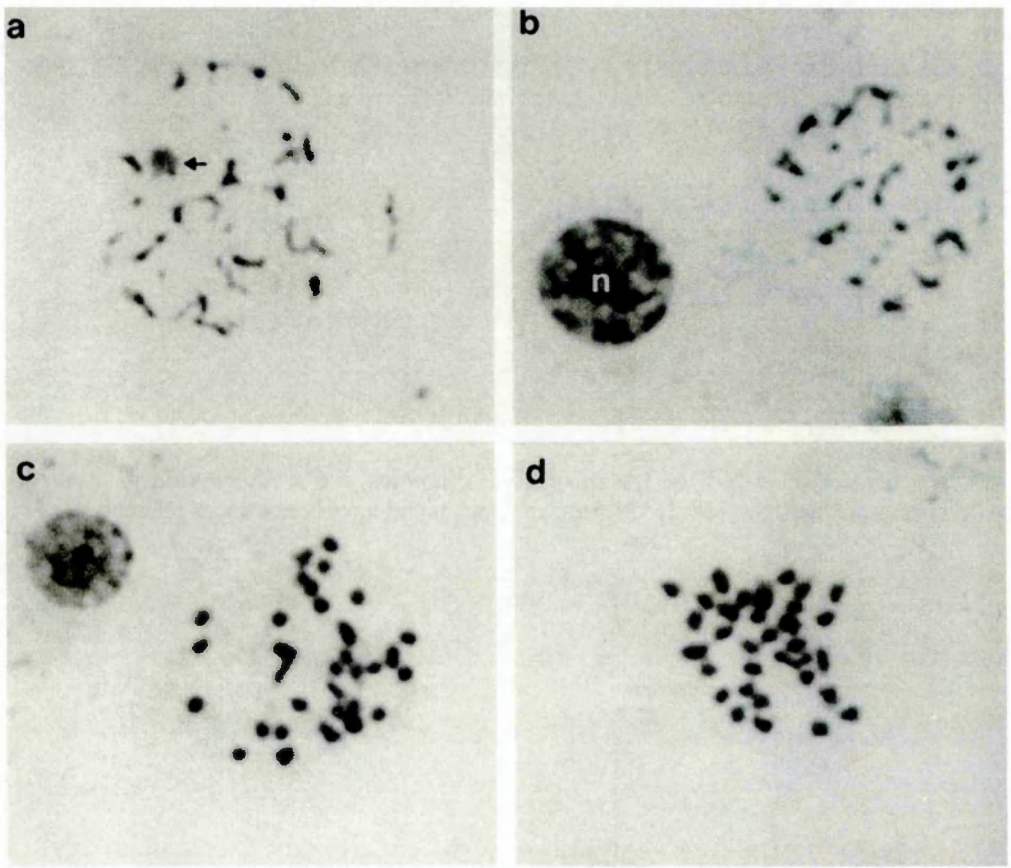


Fig. 1.9 Chromosomes of *Suberites domuncula*. The specimens have been spread after hypotonic treatment. a) prophase (the arrow points to the nucleous), b) interphase nucleous (n) on the left and prophase on the right, c) and d) condensed metaphases. The structures are visualized by bright field microscopy. Magnification x4,000. (From Imsiecke et al., 1995).

The chromosomes of the freshwater sponge *Spongilla lacustris* were visualized microscopically (Imsiecke et al., 1993). The shape and size of the chromosomes were determined and the karyotype of this sponge was established. The karyotype of a diploid cell comprises nine different chromosomes pairs, which can be subdivided into five size classes (Fig. 1.10): class1, chromosomes 1 and 2 with a length of 2.1 µm; class 2, chromosomes 3, 1.7µm; class 3, chromosome 4, 1.4µm; class 4, chromosomes 5, 1.0 µm; class 5, chromosomes 6 to 9, $\leq$ 0.7 µm. Owing to the very small size of the chromosomes it is difficult to state exactly the position of the centromeres. Chromosomes 1 and 2 were classified as metacentric, while all others seem to be telocentric. In prophase the chromosomes are arranged separately and are condensed. A large nucleolus, which is characteristic of archeocytes, is clearly visible and has a diameter of about 2.5 µm. After the disappearance of the nucleolus and the nuclear envelope, the chromosomes are arranged in the middle of the spindle apparatus along the metaphase plate. A steady increase in condensation of the chromosomes occurs during progression to metaphase. During anaphase the chromosomes separate into the corresponding sister chromatids. In telophase the chromosomes are again arranged in a compact manner.

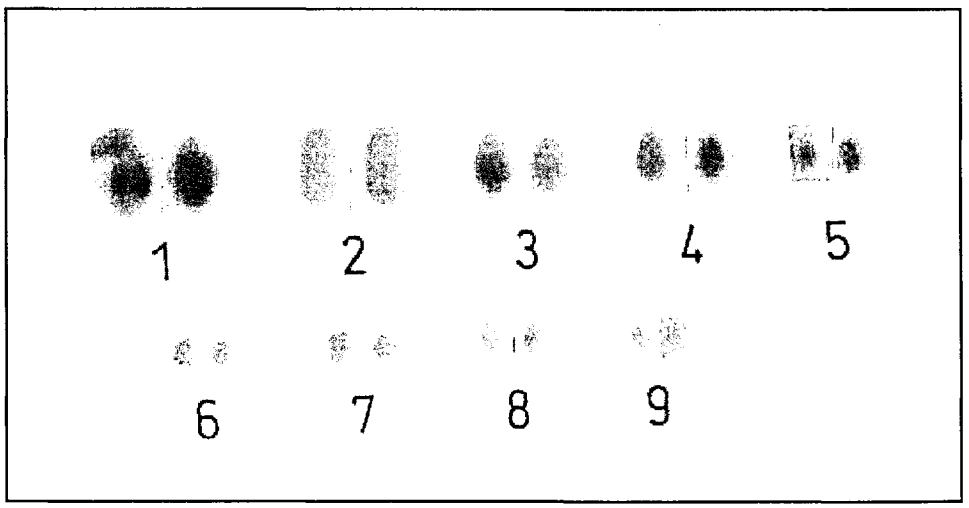


Fig. 1.10 Karyotype (diploid) of the sponge *Spongilla lacustris*. Magnification, x2900. (From Imsiecke et al., 1993).

1.3 Aim of work

The study of the genome organization in sponges is the goal of the experimental work for this research project.

Because of their basal position in the Metazoan phylogeny and of their being the simplest multicellular animals, sponges are the best system 1) to test whether the transition from unicellularity to multicellularity was accompanied by changes in the genome organization, and 2) to compare their gene distribution patterns with those of higher animals.

The first part of this investigation was devoted to the analysis of the GC level heterogeneity of the DNA in genomes of the two sponges, *Suberites domuncula* and *Geodia cydonium,* that belong to the class of Demospongiae.

Secondly the gene distribution in the genome of Demospongiae was assessed.

Because of the abundant presence of associated organisms with both sponges in analysis reported in literature, our attention was turned to the identification of these organisms, in particular Bacteria, Archaea and Algae.

Chapter 2

- Materials and Methods -

2.1 Sponge collection

The marine sponges *Suberites domuncula* (Porifera, Demospongiae, Tetractinomorpha, Hadromerida, Suberitidae) and *Geodia cydonium* (Porifera, Demospongiae, Astrophorida, Geodidae) were collected in the bay of Naples at a depth of 20 metres by the fishing service of our Institute. Individual specimens were placed separately into plastic bags and kept in seawater basins at a temperature of 15-20°C.

2.2 Extraction of genomic DNA

Genomic DNA was extracted from the internal part of the sponge body to avoid contamination of associated epibionts. Sponges were cut into small pieces and 5g of tissue was ground in liquid nitrogen and dissolved in 10 ml buffer NaCl 100 mM, EDTA 50 mM pH 8. Sodium dodecyl sulfate (SDS) solution (20%) was added to a final concentration of 2% and the mixture heated to 60°C for 30 min (Bartmann et al., 1997). Proteinase K (3 h at 50°C) and RNAse (3 h at 37°C) treatments were done. Nucleic acids were extracted with phenol/chloroform, chloroform/isoamyl alcohol and after precipitation with NaAc 3M pH 5.9 and ethanol. The DNA so extracted was dissolved in TE (10 mM Tris-HCl, EDTA 50 mM pH 8) and stored at 4°C. Genomic DNA so extracted was checked on an ethidium bromide-stained 0.7 % agarose gel (Biorad) in TBE (see Sambrook et al., 1989), visualized on GelDoc 2000 (Biorad) and quantized using a spectrophotometer UV/Vis Spectometer

Lambda Bio40 (Perkin Elmer). The DNA was analyzed also on Pulsed-Field Gel Electophoresis (PFGE) to estimate the molecular weight distribution.

Genomic DNA was also extracted from dissociated sponge cells. After washing in artificial sea water (ASW: Na_2SO_4 7 mM, $NaHCO_3$ 2 mM, Tris-HCl 20 mM, KCl 10 mM, NaCl 540 mM, $MgCl_2$ 50 mM, $CaCl_2$ 10 mM, pH 8.2), about 5 g of *Suberites domuncula* tissue was dissociated in 50 ml of calcium and magnesium-free artificial seawater containing EDTA (CMFSW-E: ASW minus $MgCl_2$ and $CaCl_2$ + 20 mM EDTA) (Müller et al., 1981) under gentle shaking at 20°C. For the silicious sponge *Geodia cydonium* the dissociation was performed in CMFSW-E supplemented with trypsin (100 µg/ml) (Müller and Zahn, 1973), penicillin (100 IU/ml) and streptomycin (100 µg/ml) (Müller et al., 1999). The cellular suspension so obtained was filtered through 20 µm mesh nylon net. The cells obtained by centrifugation at 800 x g for 15 min and after washing twice with calcium and magnesium-free artificial seawater (CMFSW: ASW minus $MgCl_2$ and $CaCl_2$) were dissolved in CMFSW. The lysis solution (4 M guanidinium thiocyanate, 25 mM sodium citrate pH7, 0.5% sarcosyl, 0.1 M 2-mercaptoethanol) was then added (0.1 ml from pellet of freshly dissociated sponge cells in 0.9 ml of lysis solution). As for DNA extraction, see above.

Genomic DNA extracted from *Geodia cydonium* was purified by equilibrium centrifugation in CsCl-Ethidium Bromide gradient (Sambrook et al., 1989).

2.3 Separation of cells

Dissociated cells were fractionated according to density via centrifugation (1000 x g for 15 min) across discontinuous Ficoll gradient centrifugation (Flowers et al., 1998; Müller et al., 1981). The Ficoll layers used were: 4%, 6%, 8%, 10%, 12.5%, 15%, 17.5%, 20%, 25%, 30% in CMFSW. The bands of cells that accumulated at the density interfaces were isolated individually by pipette, washed twice with CMFSW to remove Ficoll and pelleted at 1000 x g and 4°C for 10 min. The genomic DNA was extracted following the protocol used for dissociated sponge cells (see above).

2.4 Equilibrium centrifugation in CsCl density gradient

The profile of the DNA distribution in a CsCl gradient was obtained by analytical ultracentrifugation to sedimentation equilibrium, as previously described (Thiery et al., 1976; Sabeur et al., 1993). Standard speed was 44,000 revs/min for CsCl work using the XL-A analytical ultracentrifuge; standard wavelength was 260 nm. Concentrations of DNA should result in maximal absorbance (optical density or O.D.) between 0.3 and 1.0. 24 hours should be allowed for sedimentation equilibrium to be reached. The relationship of Schildkraut et al. (1962), $\rho = (GC \times 0.098) /100 + 1.66$, was used to convert buoyant densities into GC levels. *Bacillus subtilis* phage 2C DNA ($\rho = 1.742$ g/cm^3) was used as a density marker (Cocito, 1969).

2.5 DNA fractionation and gene distribution

DNA fractionation was performed using the "shallow gradient" method. This procedure, used first to estimate the GC content of yeast artificial chromosomes (De Sario et al., 1995), was modified for the fractionation of genomic DNA to obtain a preparative CsCl profile. Ten micrograms of DNA in CsCl + TE solution (refractive index = r.i. 1.3993) were loaded on each gradient. Centrifugation was carried out in a vertical VTi90 rotor at 20°C and 35,000 rpm for 24h, using a Beckman preparative ultracentrifuge with the brake off. About 60 fractions of 80 µl each were collected using a Hitachi DGF-U instrument. The refractive index was read for the fractions from 10 to 55 and the value of buoyant density was obtained applying the relationship

$$(10.861 \times \text{r.i.}) - 13.4974.$$

The absorbance at 260 nm of 10 µl of each fraction was measured by UV/Vis Spectometer Lambda Bio40 (Perkin Elmer) to obtain the shallow gradient profile.

The shallow gradient fractions containing the DNA were purified from CsCl with MicroSpin S-200 HR columns pre-equilibrated in TE buffer (Amersham Pharmacia Biotech Inc) following the instructions of the manufacturers. The fractions so purified were analyzed on 1% agarose gel and ethidium bromide-stained.

To assess the gene distribution, a PCR approach on the shallow gradient fractions was applied. The oligonucleotide primer sequences, used for the PCR, were designed on the basis of cDNA sequences in GenBank on TaxBrowser (Taxonomy available at www.ncbi.nih.nlm.gov). The base composition was determined using Codon W 1.3 (J. Peden; http://molbiol.ox.ac.uk/Win95.codonw.zip).

The selected primers were synthesized by the Molecular Biology Service of our Institute. The oligonucleotide primer sequences for *Suberites domuncula* and *Geodia cydonium* genes are reported in Table 2.1 and in Table 2.2.

The annealing temperature was calculated with PROLIGO – Oligos Parameter Calculation (available at www.gensetoligos.com/Calculation/calculation_frame.html).

PCR was performed using 3 ng of DNA, 25 pmol of each primer, $MgCl_2$ final concentration 2.0 mM, 10x buffer, 2 mM dNTP and 2.5 U Taq DNA Polymerase (Invitrogen). PCR was conducted on GeneAmp PCR System 9700 (Perkin Elmer). Cycling conditions were as follows: initial denaturation at 94°C, "n" cycles of 94°C for 1 min, T ann for 1 min and 72°C for 1 min ("n" = number of cycles and Tann = annealing that depend on the used primers couple, see Table 2.1 and 2.2), and a final extension of 10 min at 72°C. Each PCR product was checked by electophoresis in 1% agarose gel.

Table 2.1 Sequences of PCR oligonucleotide primers for *Suberites domuncula.*

Gene	Primers 5' - 3' (Tm)	Tann[°C]	PCR cycles
Bcl-2 homolog	**BHP1_Sd1 (f)** CGGGAGAACCTCTCATACGA (62°C) **BHP1_Sd2 (r)** CTTGATATCTGGTGCGAGTG (60°C)	58	25
Ras protein	**Ras_Sd1 (f)** GTGGTAGTCGGTGGAGGAG (62°C) **Ras_Sd2 (r)** CTGTGCTCTTCTAATGAC (52°C)	58	25
Cytochrome P450	**CytP450_Sd (f)** GACCTAGATGTAATGATG (54°C) **CytP450_Sd (r)** GATCGTCTCATCTTGGAC (54°C)	56	30
Calmodulin	**Cal_Sd1 (f)** CAAGGAGGCTTTCTCCCTCT (62°C) **Cal_Sd2 (r)** TTGCTTGTCATCATCCCAAC (62°C)	58	25
Serine/Threonine protein kinase	**cPKC_Sd3 (f)** GTGTTTCTGGCTGAGCAA (54°C) **comPKCr (r)** CCAAAGTCAGCTATCTTGA (54°C)	58	25
Glutatione peroxidase	**Gluper_Sd (f)** CATGACTGGCTTGGAGAC (56°C) **Gluper_Sd (r)** CAACTAAGTAGCACAATAC (52°C)	56	30
Polyubiquitin	**Polyu_Sd1 (f)** GCTTCTGACACCATTGAG (54°C) **Polyu_Sd2 (r)** GACGGCATACATACATAC (52°C)	54	30
Tetraspanin-CD63 receptor	**CD63R_Sd1 (f)** CGTGCGGACACTGCCTGC (62°C) **CD63R_Sd2 (r)** CGGTGAATGCAGAGACACAC (62°C)	58	25
Myol protein	**Myol_Sd (f)** GACATCGTCTGGCTAGGC (58°C) **Myol_Sd (r)** GAGAATGAGCAATAACTG (50°C)	54	30
Dermatopontin	**Der_Sd (f)** GCACTCCATGCTGTTGC (62°C) **Der_Sd (r)** CATGTGTACAGTCATAGTG (54°C)	54	35
Allograft inflammatory factor-1	**Aif_Sd (f)** CTGTGCTGTACCGATTC (52°C) **Aif_Sd (r)** GAACTAAGGCAAGTCAGC (54°C)	56	35
Cortactin	**Cor_Sd (f)** CTGATCGACTCGACTGG (54°C) **Cor_Sd (r)** GTAGCACGTACTGCAGAC (56°C)	56	45
C-jun N-terminal kinase	**Jnk_Sd (f)** CGACCGCCATAATGTCTTC (60°C) **Jnk_Sd (r)** CAGATGCACTGTTATTGTAC (56°C)	58	45
SNO protein	**SNO_Sd (f)** GTGGTCCACCTCAGATTGC (60°C) **SNO_Sd (r)** GTTGCTATGAGATGGTCCTG (60°C)	60	35
Col protein	**Col_Sd (f)** GCTGCAGTTACACTACTAG (56°C) **Col_Sd (r)** GTGCAGACAACACAGTTG (54°C)	56	35
LAGL protein	**LAGL_Sd (f)** CTCTGATCGCATATCGATC (56°C) **LAGL_Sd (r)** GCTATTGGCGCCATTGGTC (60°C)	58	45
Profilin	**Prof_Sd (f)** GCACGAGAAGTCAAGGTG (56°C) **Prof_Sd (r)** GCATTACATGCCCAGACTC (58°C)	58	45

Tm = melting temperature of the primer
Tann = annealing temperature for PCR
PCR cycles = number of cycles for PCR

Table 2.2 Sequences of PCR oligonucleotide primers for *Geodia cydonium.*

Gene	Primers 5' - 3' (Tm)	Tann [°C]	PCR Cycles
Bcl-2 homolog	BHP1_Gc1 (f) ATGGCCACTGGGTCACTGAC (64°C) BHP1_Gc2 (f) TTATCTCCCTATGATGGTCC (58°C)	58	30
Protein kinase C	cPKC_Gc1(f) TGGCAGAGCACAAGGAGT (56°C) comPKCr (r) CCAAAGTCAGCTATCTTGA (54°C)	54	30
Heat shock protein 70	HSP70_Gc (f) GGCACGACGTACTCGTGTG (62°C) HSP70_Gc (r) GTCTCTGCAGCAGTGTCTG (60°C)	60	30
Polyubiquitin	Polyu_Gc1 (f) CTCAACCGTCGAAGCCTAC (60°C) Polyu_Gc2 (r) GCTAGCCTCGACCTCTAG (58°C)	60	30
Tetraspanin_CD63 receptor	CD63_Gc (f) GTGGTCAAGTCAAGCTGC (56°C) CD63_Gc (r) GTATAGTAGAGGTCCTCG (54°C)	60	30
Thioredoxin	Thio_Gc (f) GCAGAGCGGATTCTGCCTG (76°C) Thio_Gc (r) CACTTATACATGTTGAGC (50°C)	65	30
2-5A synthetase	2-5Asyn_Gc (f) CAGAGTCTCCAGAGCTAC (56°C) 2-5Asyn_Gc (r) CTATGAACTAATCCAATG (48°C)	56	30
DNA J protein	DNAJ_Gc (f) GTACGAGGTTCTGGAGCTG (60°C) DNAJ_Gc (r) GACAAGCAGCTGCTGCC (56°C)	60	30
Leukotriene B4 protein	LB4_Gc (f) CGCAAGTACGTACTCGC (54°C) LB4_Gc (r) GCCTTCAGTGACATGTTC (54°C)	54	30
Galectin	Gal3_Gc (f) CATGGCGCGGGATTAGG (52°C) Gal3_Gc (r) CAAGCTATGCATCCAACG (54°C)	56	40
Multiadhesive protein	Muad_Gc (f) CTGGTTCTTCTGCAGGTG (56°C) Muad_Gc (r) GTAGAGTTGGAGCATACG (54°C)	56	40
Cathepsin	Cat_Gc (f) GAGCACTCAGATAGTTCC (52°C) Cat_Gc (r) GCATTGTCTGTCACGG (50°C)	56	35
Mucus-like protein	Mu_Gc (f) CAGACGACCCTCTTCAC (54°C) Mu_Gc (r) CAGCTTGTTGAGATCCATAG (58°C)	56	35
LMP7-like protein	LMP7_Gc (f) GCAGAGCATTATTCGTCGC (58°C) LMP7_Gc (r) GGGTATACAGTAGTACAG (52°C)	56	35
GDP-dissociation inhibitor	GDP_Gc (f) CATCATGGATGAGAAGTAC (54°C) GDP_Gc (r) CTCAGCTCCTCCTCGGG (58°C)	54	45
Beta-gamma-crystallin	Cry_Gc (f) CAGCAGCACTGAACTCCC (58°C) Cry_Gc (r) GTAAACTCTCTAGCTAGC (52°C)	58	45
Tubulin	Tub_Gc (f) CAGTGCGGCAACCAGATTG (60°C) Tub_Gc (r) GCTCTCCCTCCTCACACC (60°C)	62	45
Rh antigen-like protein	Rh_Gc (f) CAGGATTTCTGCTGGTGTTC (60°C) Rh_Gc (r) CAGCACTGCGGCCATCTC (60°C)	62	45

Tm = melting temperature of the primer
Tann = annealing temperature for PCR
PCR cycles = number of cycles for PCR

2.6 Amplification, cloning and sequencing of eukaryotic 5.8S-28S rDNA, prokaryotic 16S rDNA and Archaea 16S rDNA

The amplification of eukaryotic 5.8S-28S rDNA was done with universal eukaryotic primers ITS3-D2 (Christen et al., 1991; Lafay et al., 1992), that of prokaryotic 16S rDNA with primers 27F-1385R (Grigioni et al., 1999), that of Archaea 16S rDNA with archaea specific-primers Ar4F-1119aR (Jurgensen et al., 2000) (Table 2.3). A 25 ng aliquot of DNA was amplified. PCR was performed using 25 pmol of each primer, $MgCl_2$ final concentration 2.0 mM, 10x buffer, 2 mM dNTP and 2.5 U Expand High Fidelity PCR System (Roche). PCR was done on GeneAmp PCR System 9700 (Perkin Elmer). Cycling conditions were as follows: initial denaturation at 94°C, "n" cycles of 94°C for 1 min, T ann for 1 min and 72°C for 1 min.

PCR products were analyzed by electrophoresis in 1% agarose gel. Purified PCR products (QIAquick PCR Purification Kit, Quiagen) were cloned into the pCR 2.1 plasmid vector and transformed into *E. coli* competent cells using the commercial kit Original TA Cloning (Invitrogen) following the instructions of the manufactures. Plasmid DNA was extracted using QiAprep Spin Miniprep Kit (Qiagen) and inserts were sequenced in a CEQ 2000 Beckman automatic sequencer by the Molecular Biology Service of our Institute.

Sequences were compared to those in databases using the Basic Local Alignment Search Tool (BLAST, Altschul et al., 1997) algorithm (available at www.ncbi.nih.nlm.gov) to identify known sequences with a high degree of similarity. The alignments between the sequences were done using MultAlin (available at prodes.toulouse.inra.fr/multalin/multalinl.html). Evolutionary trees were generated using

maximum parsimony algorithms in the PHYLIP package (version 3.4; J. Felsenstein, University of Washington, Seattle).

Table 2.3 Sequences of the oligonucleotide primers used for PCR.

	Primer 5'-3' (Tm)	T ann [°C]
Eukaryotic 5.8-28S rDNA	**ITS3** GTCGATGAAGAACGCAGC **D2** TCCGTGTTCAAGACGGG	60
Prokaryotic 16S rDNA rDNA	**27F** GAGTTTGATCCTGGCTCAG **1385R** GGGTGTGTRCAAGGCCC	55
Archaea 16S rDNA	**Ar4F** TCYGGTTGATCCTGCCRG **1119aR** GGYRSGGGTCTCGCTCGTT	60

Chapter 3

- Results and discussion -

3.1 Heterogeneity of the base composition in sponge DNA

Before presenting the experimental work, it is relevant to give a brief introduction on the two sponges analyzed. Figs. 3.1 and 3.2 show *Suberites domuncula* and *Geodia cydonium*, respectively: both live in the sea of Naples. *Suberites domuncula* lives in the Gulf of Mergellina and Posillipo in Naples in a depth range from 14 to 16 metres. The body of *Suberites domuncula* (Fig. 3.1) has an orifice in which lives a hermit crab *Pagurites oculatus* (*Decapoda*: *Paguridea*), which resides inside shells of the mollusc *Trunculariopsis trunculus* (emerging in Fig. 3.1b). Because of the presence of this hermit crab, *Suberites domuncula* has the possibility to move.

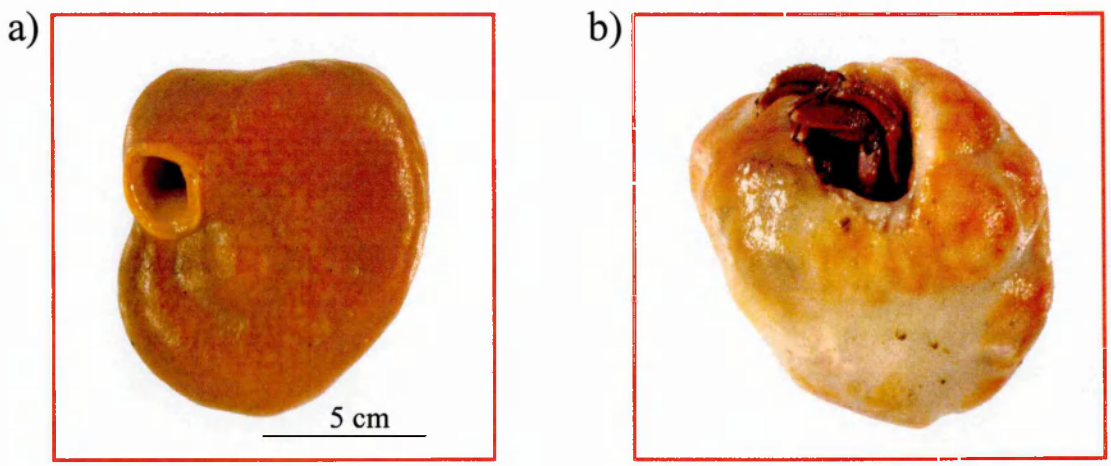


Fig. 3.1 Photo of *Suberites domuncula* (a) The part in red is the body of *Suberites domuncula* that has an orifice in which lives a hermit crab *Pagurites oculatus* (*Decapoda: Paguridea*), which resides inside shells of the mollusc *Trunculariopsis trunculus* and emerging in b).

In contrast, *Geodia cydonium* lives in the Gulf of Bacoli and Baia, near Naples, in a depth range from 2-3 to 15 metres, on the sandy seabed and covered with mud. In fact, the surface of *Geodia cydonium* is always very dirty (see Fig. 3.2).

Fig. 3.2 Photo of *Geodia cydonium.*

The seawater around both sponges has an average temperature of about 20°C.

It should be stressed that is very problematic to isolate pure sponges DNA, due to the associated bacterial and eukaryotic organisms which cannot be easily separated from the sponge tissues.

Genomic DNA was extracted from tissue of *Geodia cydonium* and *Suberites domuncula* and analysed by analytical ultracentrifugation. Fig. 3.3 shows the CsCl analytical ultracentrifugation profile of genomic DNA from *Geodia cydonium*.

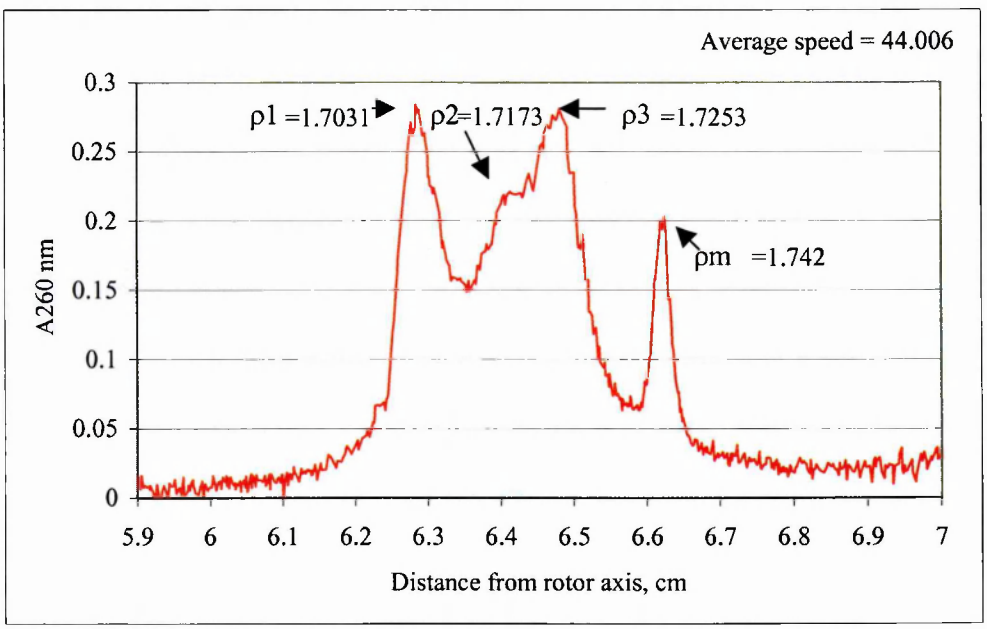


Fig. 3.3 Profile of G*eodia cydonium*. DNA extracted from whole tissue as obtained by analytical ultracentrifugation to sedimentation equilibrium in a CsCl gradient. Bacteriophage 2C is used as a marker (ρ = 1.742). Density values are in g/cm^3. Experimental error of density values is 0.0005.

Three peaks are visible and characterized by different values of buoyant density (ρ1= 1.7031 g/cm^3, ρ2= 1.7173 g/cm^3, ρ3= 1.7253 g/cm^3). Previous analysis suggested that *Geodia cydonium* DNA is very heterogeneous (Bartmann et al., 1997). The authors claimed

that the profile could be described satisfactorily by the superposition of at least five components (Fig. 3.4), whose buoyant densities were 1.6972, 1.7054, 1.7128, 1.7195, 1.7262 g/cm^3, respectively. The proportion of total DNA of these components were 8%, 16%, 12%, 30%, 34%, respectively.

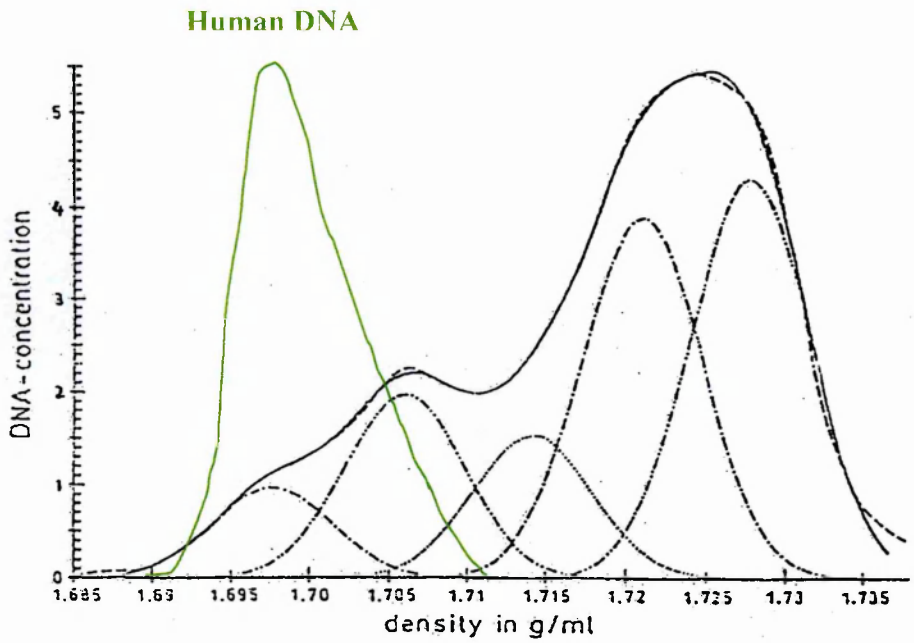


Fig. 3.4 Analytical density gradient centrifugation profile of total *Geodia cydonium* DNA. The curves represent: the measured profile (dashed line), the subcomponents, obtained from curve fit calculations (dashed-dotted lines), the profile from the sum of subcomponents (solid line). The human DNA profile is shown in green. (Modified from Bartmann et al., 1997).

Bartmann et al. (1997) excluded bacterial contamination of *Geodia cydonium* DNA based on the reassociation constants and genetic complexity of the five fractions as determined by reassociation kinetics. However, it was not possible to exclude contamination from other

eukaryotic organisms. Such an extreme heterogeneity of sponge DNA base composition, reported by Bartmann et al. (1997), is very puzzling since it has never been observed before for any organisms. Indeed, for example *Geodia cydonium* DNA would be more heterogeneous than human DNA (Fig. 3.4): the green profile in the fig. represents CsCl analytical ultracentrifugation profile for human DNA.

Fig. 3.5 shows the CsCl analytical ultracentrifugation profile of genomic DNA extracted from *Suberites domuncula*.

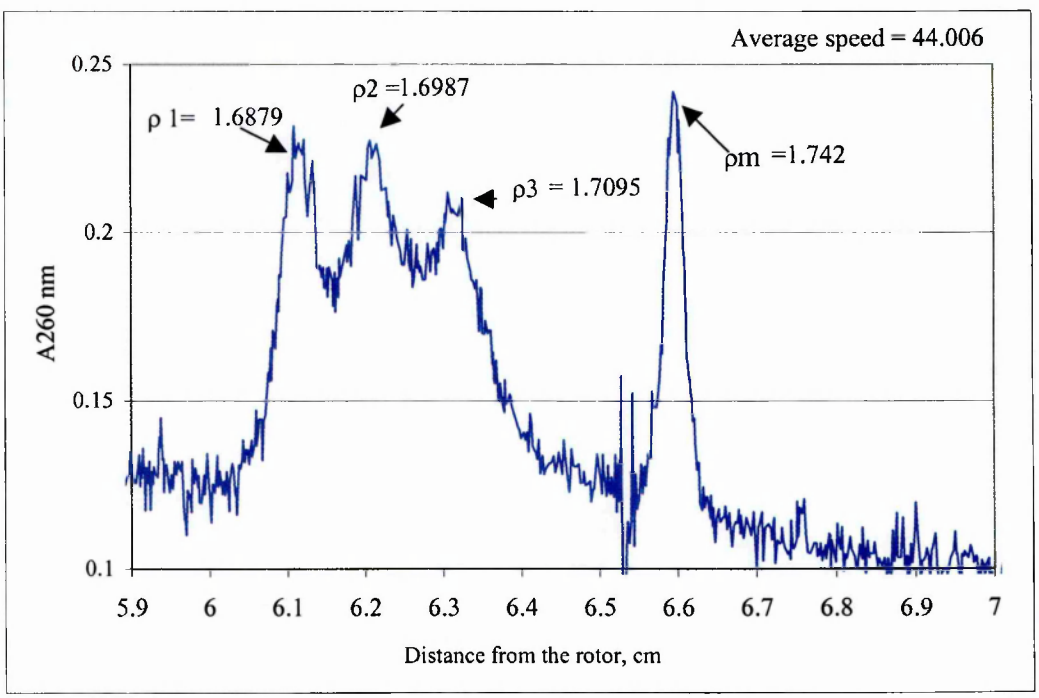


Fig. 3.5 Profile of *Suberites domuncula* DNA extracted from whole tissue as obtained by analytical ultracentrifugation to sedimentation equilibrium in a CsCl gradient.

This DNA also exhibits three peaks (ρ1= 1.6879 g/cm^3, ρ2= 1.6987 g/cm^3, ρ3= 1.7095 g/cm^3) characterized by densities different from those found in *Geodia cydonium* DNA. This would suggest that the associated organisms are different in the two *Demospongiae* species.

Two explanations can account for the presence of the three peaks in two sponge DNAs:

1) these sponge DNAs are very heterogeneous as suggested by Bartmann et al. (1997);
2) only one peak is due to sponge DNAs and the other two peaks are from associated organisms, known from the literature that are present in these two sponges.

In order to address this issue, we attempted to purify sponge genomic DNA and to identify the potentially associated organisms.

3.2 Identification of sponge DNA

Concerning the identification of sponge DNA it was possible to obtain a partial purification by the dissociation of the sponge tissue.

For this purpose, the two sponges were cut into pieces, eliminating the external layer, and put into a basin with filtered water and kept in the dark to avoid the presence of bacteria and photosynthetic organisms. This treatment lasted for about four days. The tissue so treated was dissociated (see Materials and Methods) and DNA extracted analysed on CsCl analytical ultracentrifugation.

The CsCl analytical ultracentrifugation profile obtained for *Suberites domuncula* DNA is reported in the Fig. 3.6.

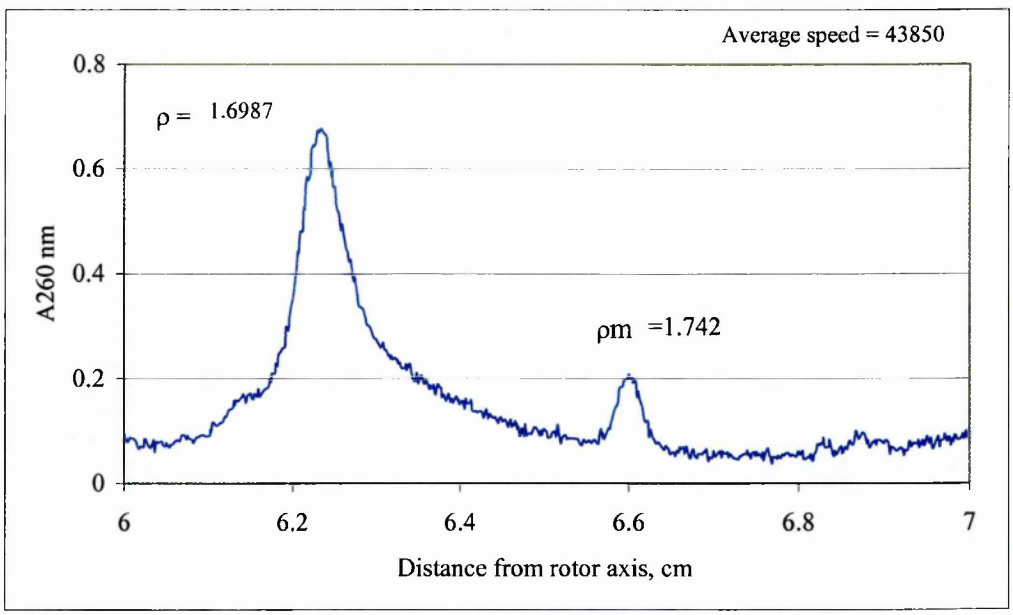


Fig. 3.6 Profile of *Suberites domuncula* DNA extracted from dissociated cells as obtained by analytical ultracentrifugation to sedimentation equilibrium in a CsCl gradient.

The single peak observed corresponds to a density value of 1.6987 g/cm^3 which corresponds to the second peak reported in Fig. 3.5. The other two peaks were almost completely eliminated (see below), and are not visible in the CsCl analytical ultracentrifugation profile.

Fig. 3.7 shows the CsCl analytical ultracentrifugation profile of *Geodia cydonium* DNA characterized by a main peak with a buoyant density of 1.7031 g/cm^3, which corresponds to the first peak reported in Fig. 3.3. The two other peaks found in the previous experiment (Fig. 3.3) were reduced in amounts.

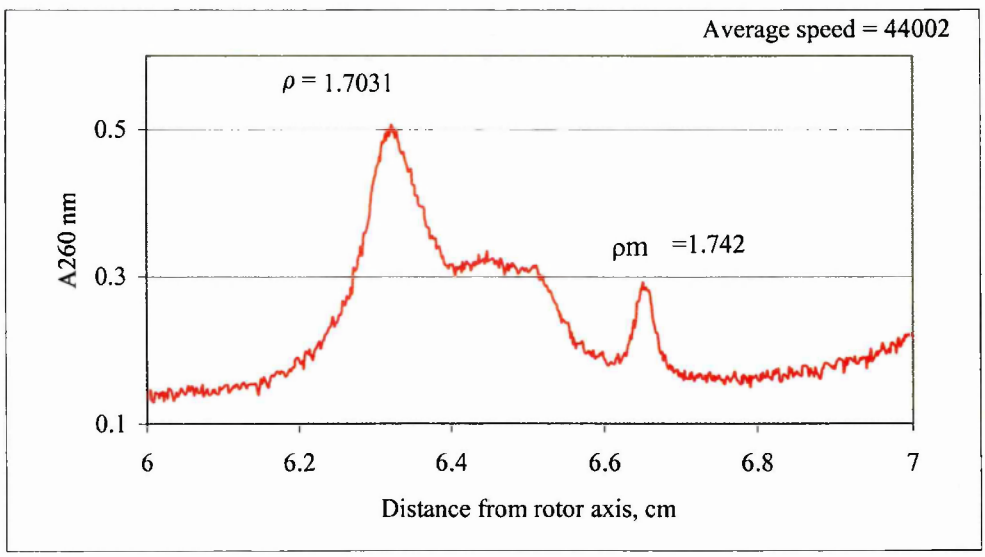


Fig. 3.7 Profile of *Geodia cydonium* DNA extracted from dissociated cells as obtained by analytical ultracentrifugation to sedimentation equilibrium in a CsCl gradient.

To proceed further in DNA purification, the sponge tissue was dissociated (see Materials and Methods). Figs. 3.8 and 3.9 display photos for *Geodia cydonium* and *Suberites domuncula* dissociated cells, respectively: in both cases different cellular types are present. Indeed, cells are different in dimensions. In *Suberites domuncula* granular cells are present, in *Geodia cydonium* are still present bacteria.

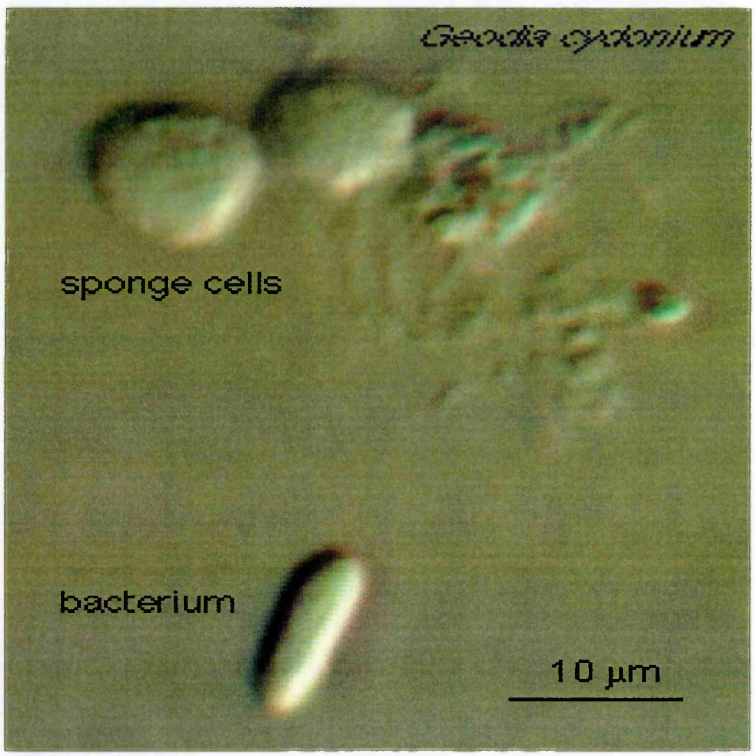


Fig. 3.8 Light microscopy picture of *Geodia cydonium* cells showing large cells and bacteria.

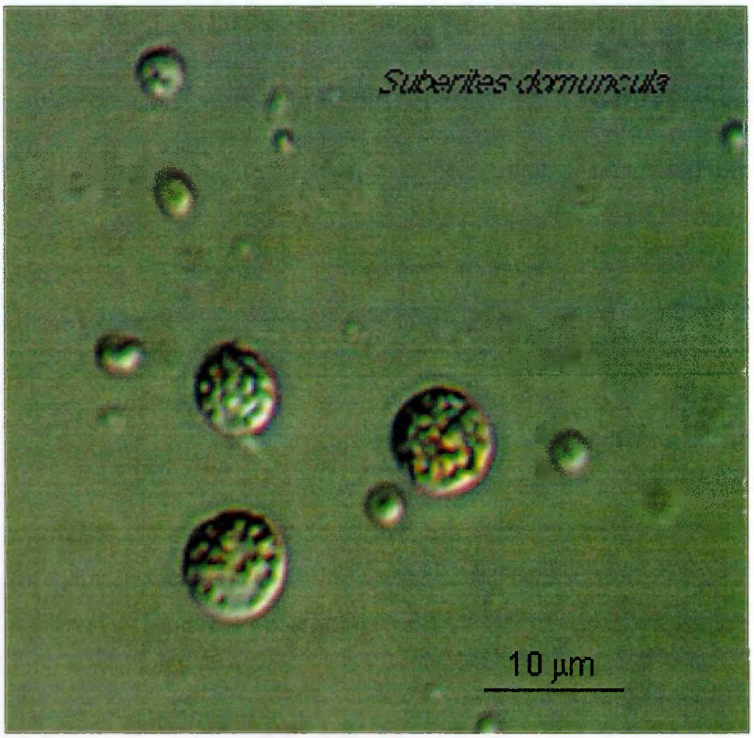


Fig. 3.9 Light microscopy picture of *Suberites domuncula* cells: as it is visible there is the presence of granular cells.

Dissociated cells from both sponges were loaded on Ficoll discontinuous gradient. Fig. 3.10 presents a scheme of cell fractionation for the two sponges. Eight cell layers (red layers) were obtained for *Suberites domuncula*, whereas five cell layers (blue layers) for *Geodia cydonium*. Microscopic analysis of each cell layers obtained showed again the presence of bacteria, suggesting that they are associated with *Geodia cydonium* and

Suberites domuncula (see below). Genomic DNA was extracted from each of these cell layers and analyzed by analytical ultracentrifugation. The profiles so obtained showed the same peaks reported in Figs. 3.6 and 3.7.

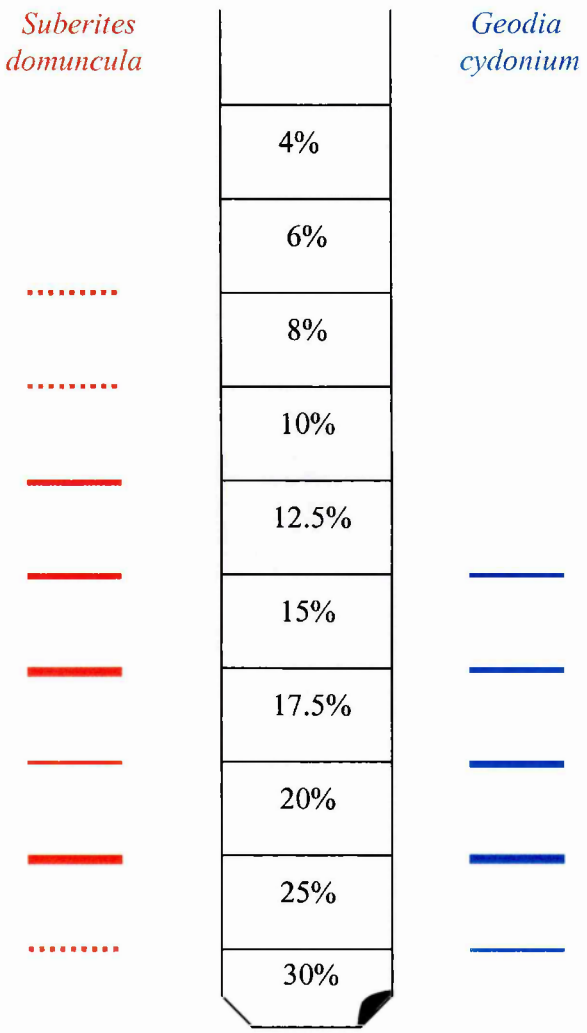


Fig. 3.10 Cell fractionation in Ficoll discontinuous density gradient. Layers of dissociated cells of *Suberites domuncula* (red cell layers) and of *Geodia cydonium* (blue cell layers) are schematically drawn. (Modified from Müller et al., 1981).

To obtain an even further purified DNA, the *Geodia cydonium* DNA was centrifugated in CsCl-Ethidium bromide gradient (see Materials and Methods). Fig. 3.11 shows the CsCl analytical ultracentrifugation profile of *Geodia cydonium* DNA obtained after this experiment: the single peak observed corresponds to the predominant peak (ρ= 1.7030 g/cm^3) found previously (Fig. 3.7) and the other two peaks (Fig. 3.3) were eliminated even if not completely, however they are not visible in the CsCl analytical ultracentrifugation profile (see below).

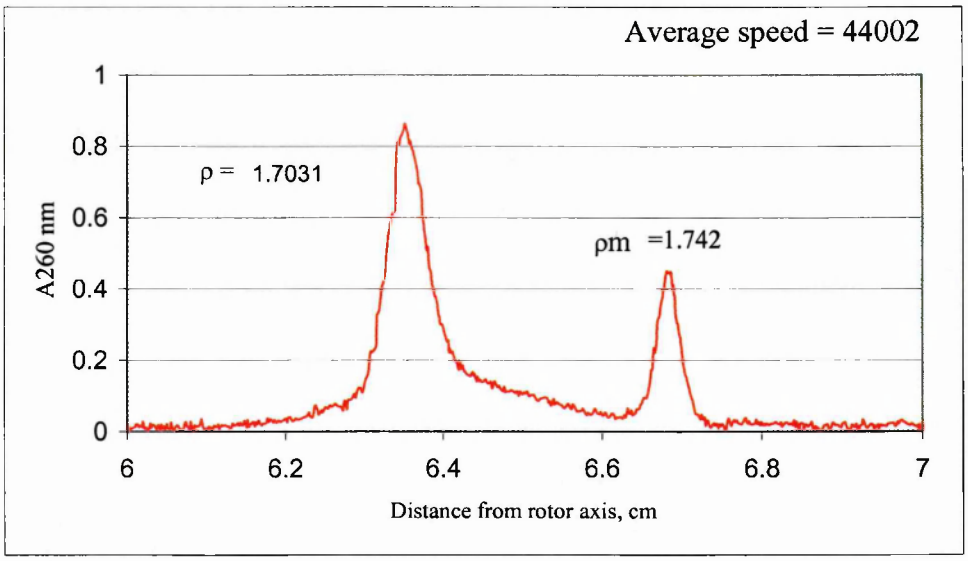


Fig. 3.11 Analytical ultracentrifigation profile of *Geodia cydonium* DNA extracted from dissociated cells after purification by equilibrium centrifugation in CsCl-Ethidium bromide gradient: the single peak found corresponds to the predominant peak (ρ= 1.7030 g/cm^3) found previously.

Fig. 3.12 shows the CsCl analytical ultracentrifugation profile of *Geodia cydonium* DNA in comparison with the DNA of *Suberites domuncula*.

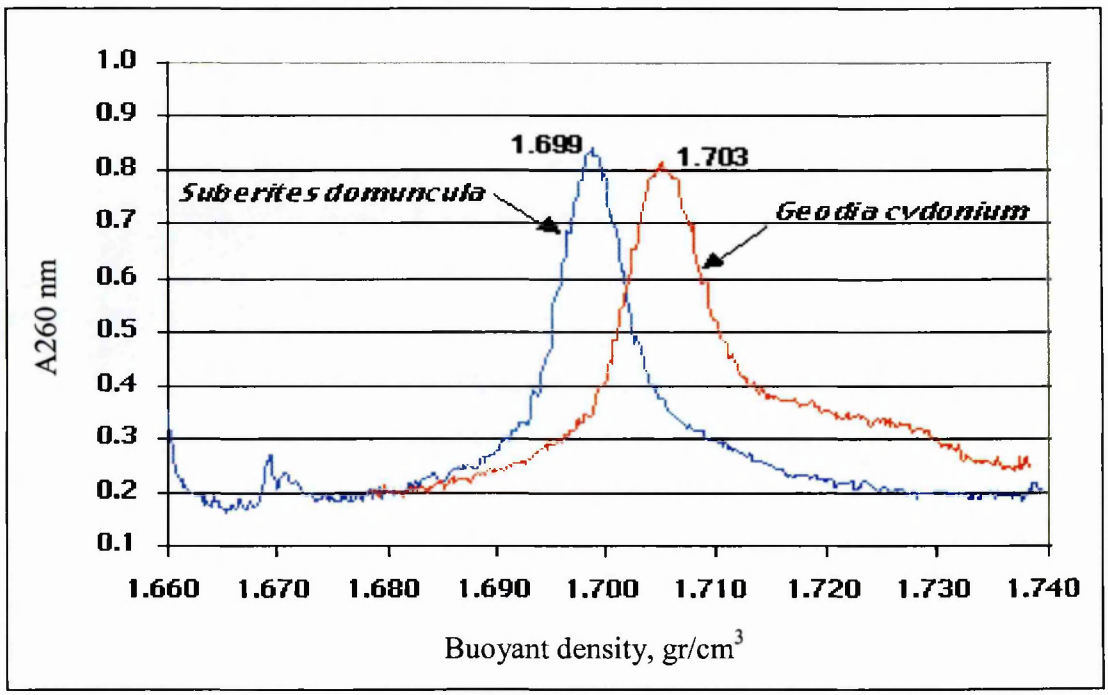


Fig. 3.12 Comparison of CsCl analytical ultracentrifugation profiles of *Geodia cydonium* and *Suberites domuncula* DNAs.

The Bartmann et al. (1997) profile for *Geodia cydonium* DNA has been reported in Fig. 3.13 for comparison with the range of heterogeneity found in this work.

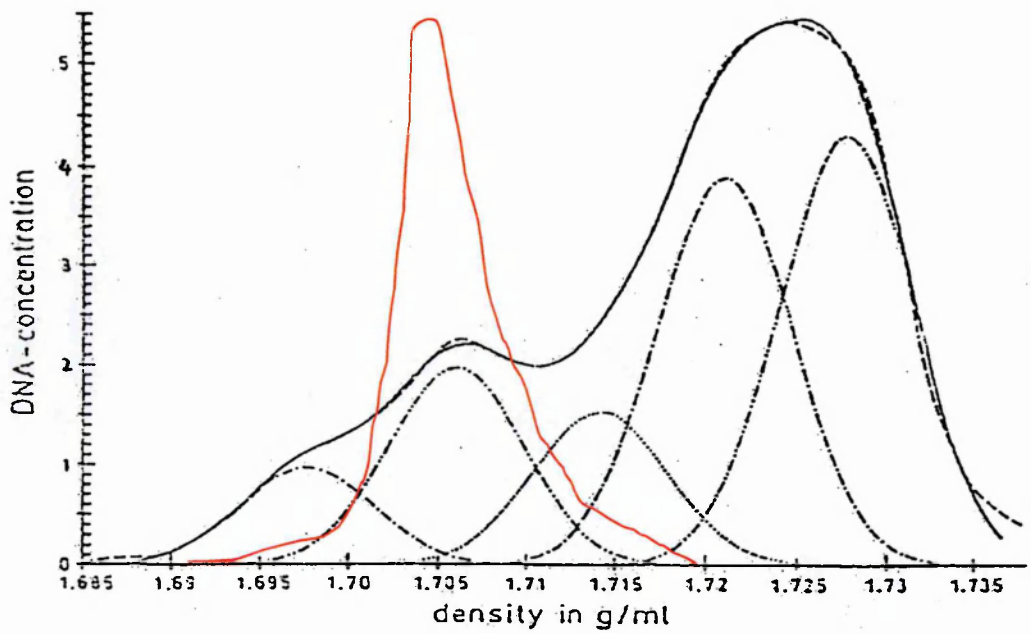


Fig. 3.13 Bartmann's profile for *Geodia cydonium* DNA in comparison with the CsCl analytical ultracentrifugation profile (in red) found in this work. (Modified from Bartmann et al., 1997).

Indeed, Fig. 3.14 shows the analytical profile of *Geodia cydonium* DNA in comparison with human DNA and *Xenopus laevis* profiles just to compare their range of heterogeneity.

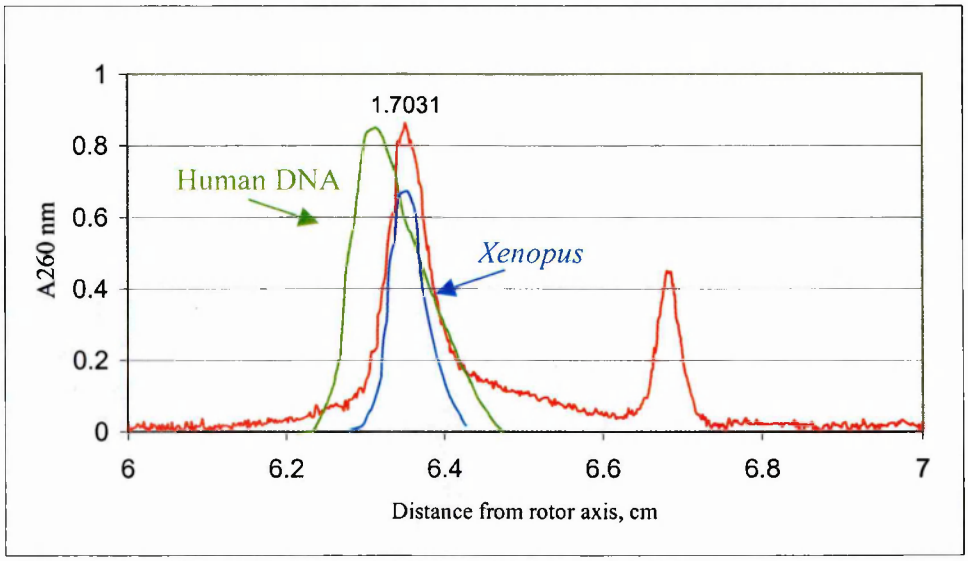


Fig. 3.14 Analytical profile of *Geodia cydonium* DNA in comparison with human DNA (green profile) and *Xenopus laevis* (blue profile) profiles.

These results indicate that the profile of *Geodia cydonium* DNA, reported by Bartmann et al. (1997), was not corresponding to sponge DNA. Probably only one was the peak due to *Geodia cydonium* DNA and whereas the other peaks were due to the presence of associated organisms that could not be eliminated from sponge DNA (see below). Probably this problem was due to the method used to extract the DNA. In fact the genomic DNA was extracted from total tissue without the type of treatment carried out in the current study.

Since a brownian diffusion was observed in the CsCl analytical ultracentrifugation profile for both sponge DNAs we determined the molecular weight of both DNAs to understand and explain their CsCl analytical ultracentrifugation profile. Both sponge DNAs were analysed by ethidium bromide gel electrophoresis: as it is possible to see in the Fig. 3.15 the molecular weight of the two DNA is about the same as Lambda (λ) DNA (48.5 kb), used as a marker but there are DNA fragments of low molecular weight.

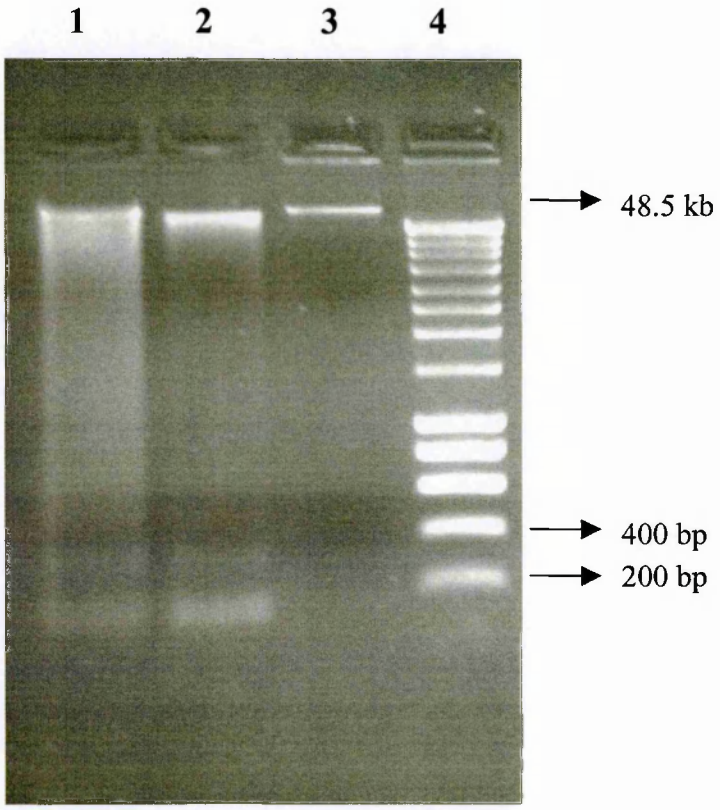


Fig. 3.15 Image of *Suberites domuncula* and *Geodia cydonium* genomic DNA observed on an ethidium bromide-stained 0.7% agarose gel.
Lane 1 = *Suberites domuncula* genomic DNA
Lane 2 = *Geodia cydonium* genomic DNA
Lane 3 = λ DNA (used as molecular weight marker)
Lane 4 = SmartLadder, molecular weight marker (Eurogentec)

Since this was not an occasional event but occur in each extraction, we thought that these fragments were due to an endonuclease activity of the sampled species.

An analysis on pulsed-field gel electrophoresis (PFGE) was also done for both DNA: in this case the range of the fragments is between 48.5 kb and 23.1 (Fig. 3.16). According to these results, the molecular weight of these sponge genomic DNAs is not so low as to justify the observed diffusion, which is probably due to the presence of the associated organisms (see below).

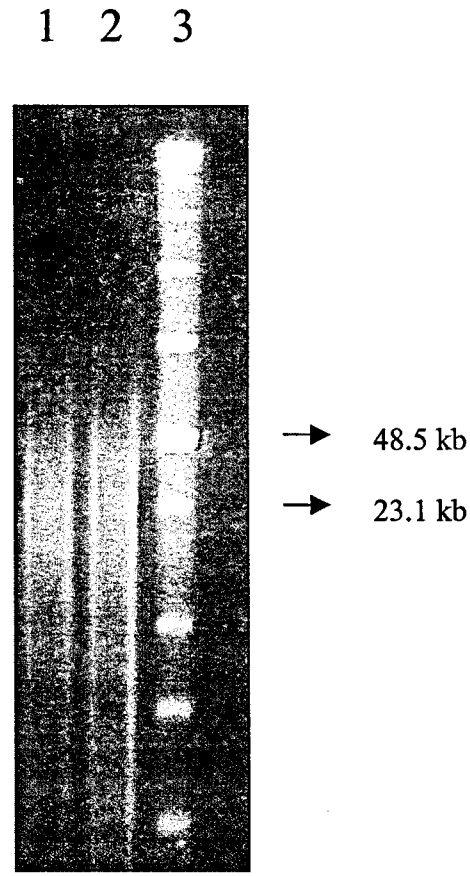


Fig. 3.16 Analysis of *Suberites domuncula* and *Geodia cydonium* genomic DNA on pulsed-field electophoresis (PFGE).
Lane 1 = *Suberites domuncula* genomic DNA
Lane 2 = *Geodia cydonium* genomic DNA
Lane 3 = Low Range PFG Marker (Biolabs)

From the buoyant density of the CsCl analytical profile for the two genomic sponge DNA, so extracted, it has been possible to calculate the GC% of both DNA, using the equation of Schildkraut et al. (1962). The GC% corresponds to 39.6 for S*uberites domuncula* DNA and 43.9 for *Geodia cydonium* DNA.

3.3 Gene distribution

The second part of this investigation was devoted to assessing the gene distribution in the genomes of *Geodia cydonium* and *Suberites domuncula*. The first step was the fractionation of DNA. The base composition heterogeneity of sponge DNA allows this DNA to be fractionated by CsCl density gradient centrifugation, using the "shallow gradient" technique (see Materials and Methods). This approach was originally developed to estimate the G+C content of yeast artificial chromosomes and then modified for the fractionation of genomic DNA. Fig. 3.17 shows the fractionation for *Geodia cydonium* DNA: 19 fractions were obtained, characterized by different buoyant densities (i.e. GC content).

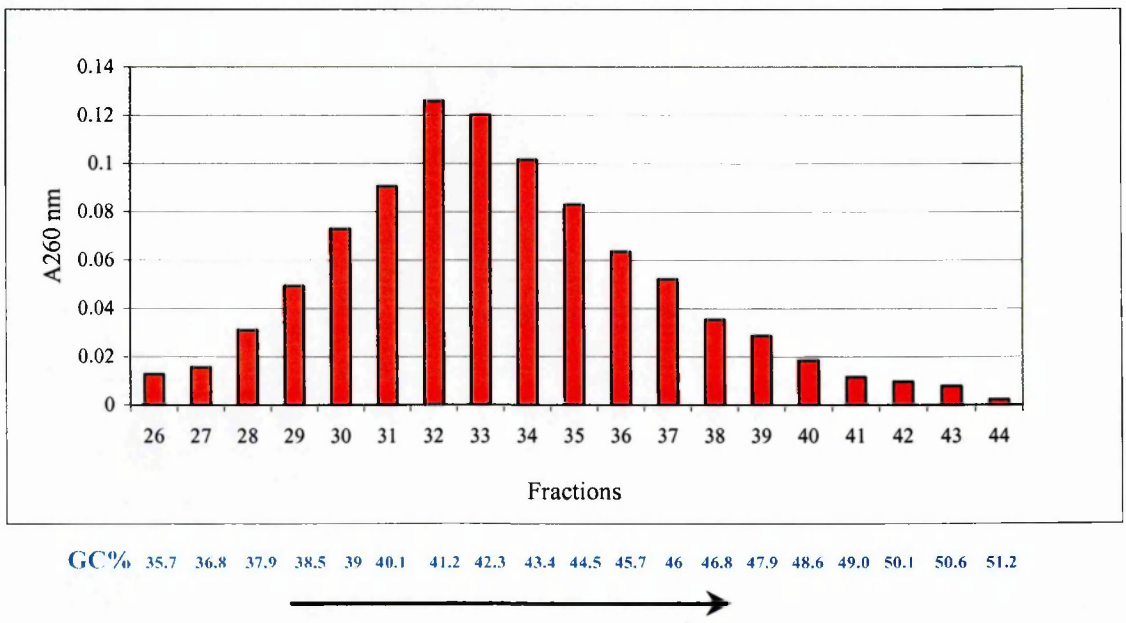


Fig. 3.17 DNA profile of *Geodia cydonium* using the shallow gradient method. Ten micrograms of genomic DNA were loaded. Numbers in blue represent the GC content (GC%) of each fraction.

Fig. 3.18 shows the fractionation for *Suberites domuncula* DNA: 25 fractions were obtained. In the two graphs the GC level increases from left to right. The modal buoyant densities of the two sponges' DNA, as obtained from shallow gradient fractionations, match those obtained by analytical centrifugation.

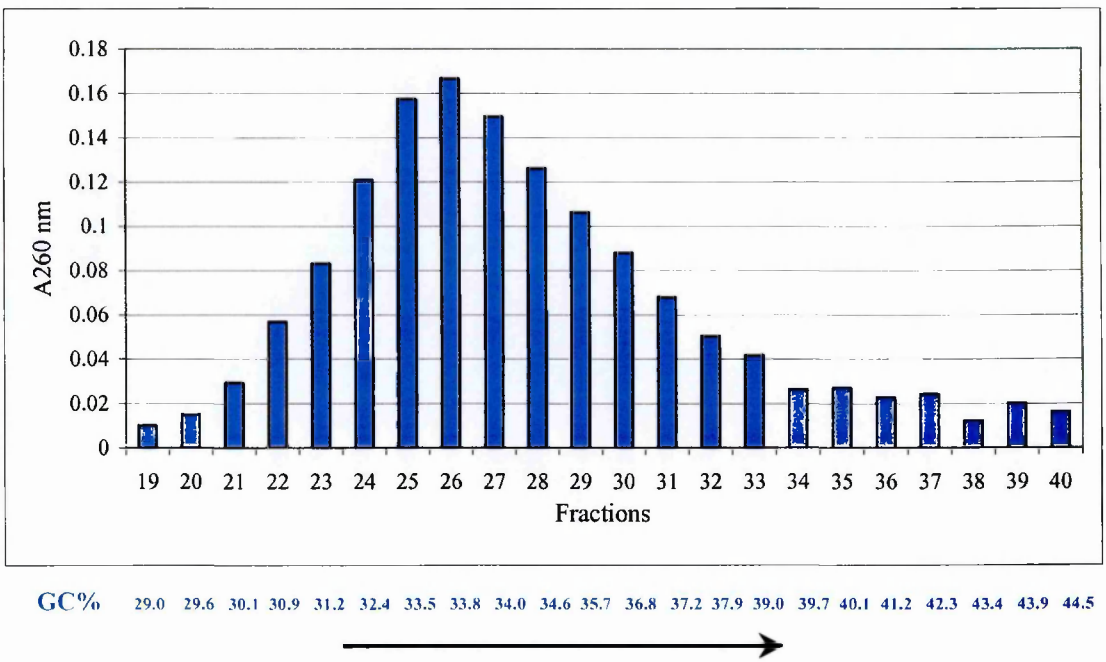


Fig. 3.18 DNA profile of *Suberites domuncula* using the shallow gradient method. The numbers in blue represent the GC content (GC%) of each fraction.

The following step was to analyse the gene sequences available in GenBank for the sponges.

The number of sponge genes in GenBank is very small: for the Demospongiae class, 57 coding sequences (cDNA or CDS) are available for *Suberites domuncula* and 78 for *Geodia cydonium*, 34 for *Ephydatia fluviatilis* (a freshwater sponge); only 8 sequences can be found for *Sycon raphanus* belonging to the Calcarea class; no cDNA sequences exist for the Hexactinellidae. Genomic DNA sequences were available only for the Demospongiae. Even if the number of genes is small, the genes available for *Suberites domuncula* and *Geodia cydonium* should have been sufficient to provide preliminary information on the gene distribution, since they cover a wide range of GC contents in third codon positions: 32-60% for *Suberites domuncula* and 28-68% for *Geodia cydonium*. For the sake of comparison, the range of GC contents in third codon positions for human DNA covers 30-95% and for *Xenopus laevis* 21-86%.

PCR amplification with specific primers used to localize genes of interest in DNA fractions.

Fig. shows an example of localization for the *Geodia cydonium* gene Hsp70: this gene was centered in fraction 30 of the shallow gradient.

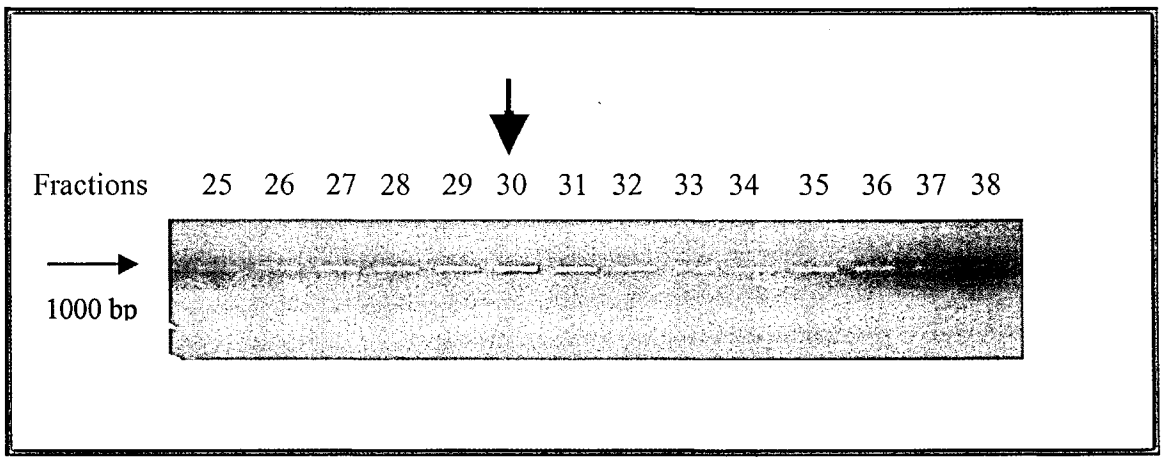


Fig. 3.19 Image of an example of localization for the *Geodia cydonium* gene Hsp70 observed on an ethidium bromide-stained 0.7% agarose gel: the gene is localized on the shallow gradient fraction 30 (blue arrow).

PCR conditions were optimized for 17 genes of *Suberites domuncula* and for 18 of *Geodia cydonium*, chosen according their GC_3 values so as to cover the distribution range of all available coding sequences of these two sponges. Tables 3.1 and 3.2 list the analysed genes for *Suberites domuncula* and for *Geodia cydonium*, with their accession numbers, lengths in amino acids, total GC% and GC_3 levels were reported respectively. Each gene reported in the table was localized on the shallow gradient fractions.

Table 3.1 Accession number, length in amino acids, GC%, GC_3%, localization on shallow gradient fractions (with GC%) of the 17 coding sequences for *Suberites domuncula*.

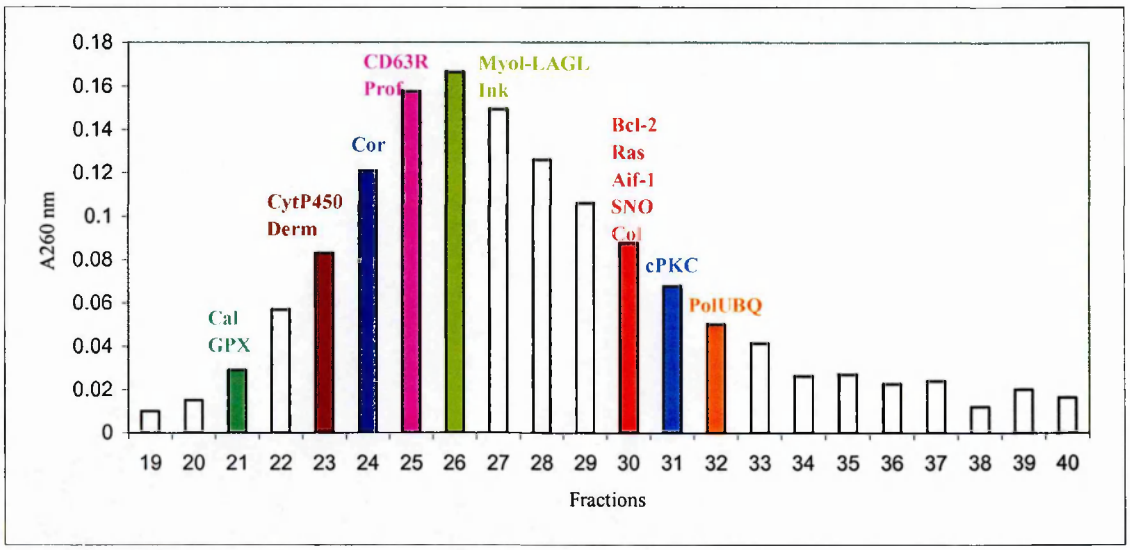
Gene	Accession No.	Length (aa)	GC, % CDS	GC3, %	Fraction	
					#	GC%
BHP1 protein	Y19158	219	41.9	39.7	30	36.8
Ras protein	Y18167	192	42.7	45.3	30	36.8
Cytochrome P450	Y17816	482	45.4	46.3	23	31.2
Calmodulin	Y18166	150	46.2	48	21	30.1
Serine/Threonine protein kinase	Y13099	674	47.2	51.9	31	37.2
Glutatione peroxidase	Y18438	218	49.1	55.0	21	30.1
Polyubiquitin	Y12081	381	49.5	55.0	32	37.9
Tetraspanin-CD63 receptor	Y18100	249	50.1	57.8	25	33.5
Myol protein	AJ252240	121	44.0	38.8	26	33.8
Dermatopontin	AJ299722	185	43.4	50.2	23	31.2
Allograft inflammatory factor-1	AJ410885	145	41.1	47.5	30	36.8
Cortactin	Y18027	478	45.8	35.1	24	32.4
C-jun N-terminal kinase	AJ291511	362	45.2	49.2	26	33.8
SNO protein	AJ277954	234	45.3	41.9	30	36.8
Col protein	AJ252241	283	48.8	28.6	30	36.8
LAGL protein	AJ250580	331	44.8	50.7	26	33.8
Profilin	Y18900	141	46.8	38.3	25	33.5

Table 3.2 Accession number, length in amino acids, GC%, GC_3%, localization on shallow gradient fractions (with GC%) of the 18 coding sequences for *Geodia cydonium*.

Gene	Accession No.	Length (aa)	GC, % CDS	GC3, %	Fraction	
					#	GC%
BHP1 protein	Y19157	256	54.2	59.0	29	38.5
Protein kinase C	Y17882	678	53.2	64.3	30	39.0
Heat shock protein 70	X94985	664	54.8	68.2	30	39.0
Polyubiquitin	X70917	458	54.9	71.6	30	39.0
Tetraspanin_CD63 receptor	Y19156	256	54.2	58.9	33	42.3
Thioredoxin	Y17147	107	53.6	77.6	26	35.7
2-5A synthetase	Y18497	328	42.3	37.8	26	35.7
DNA J protein	Y09037	413	54.2	59.6	29	38.5
Leukotriene B4 protein	Y19102	336	47.2	48.2	30	39.0
Galectin	X93925	191	44.3	38.7	38	46.8
Multiadhesive protein	Y14243	702	49.1	49.4	38	46.8
Cathepsin	Y10527	323	53.7	64.4	39	47.9
Mucus-like protein	AJ299721	539	45.9	39.7	31	40.1
LMP7-like protein	X97728	281	55	64.4	31	40.1
GDP-dissociation inhibitor	X94983	449	47.0	50.5	38	46.8
Beta-gamma crystallin	Y08771	164	49.0	51.8	35	44.5
Tubulin	Y17002	450	54	66.2	38	46.8
Rh antigen-like protein	Y12397	524	52.2	60.7	35	44.5

Figs. 3.20a) and b) shows the localization of the genes on the shallow gradient fractions.

a)



b)

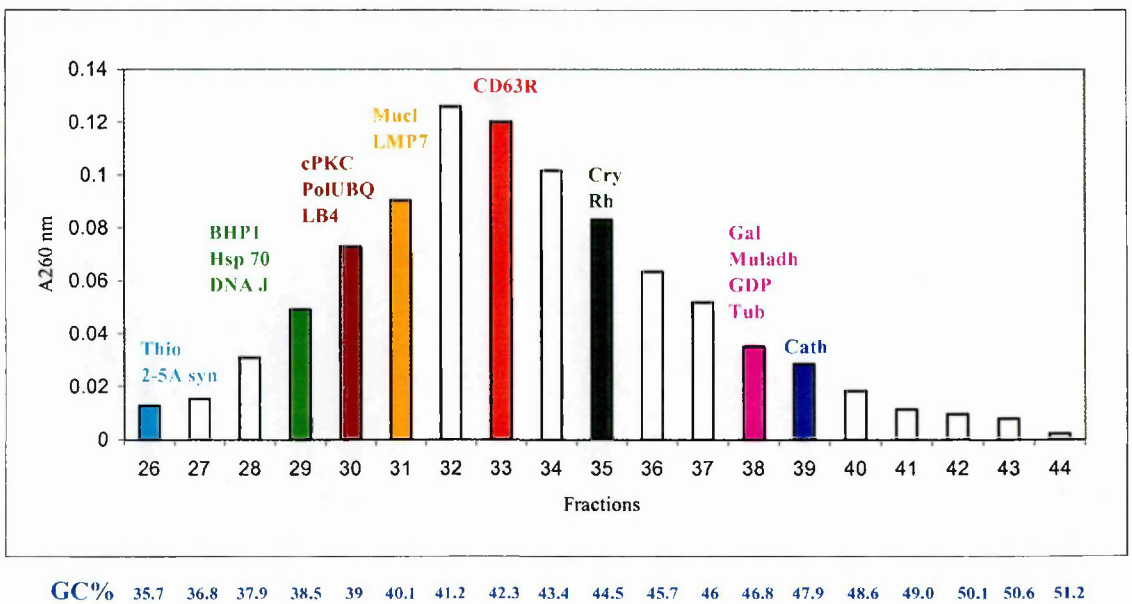
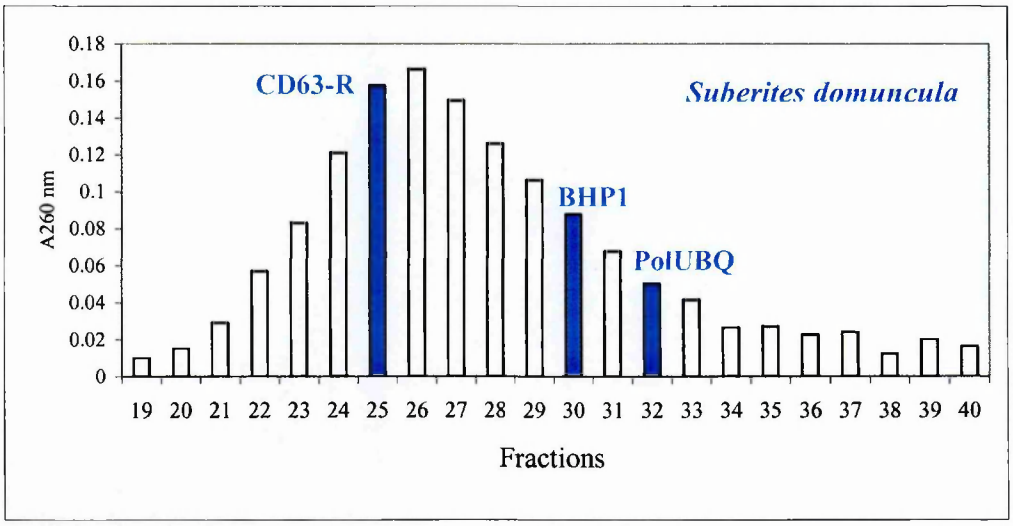


Fig. 3.20 Localization of the genes on a) *Suberites domuncula* and b) *Geodia cydonium* shallow gradient fractions. The GC% of the fractions is also shown.

The localization of the analysed coding sequences from both *Suberites domuncula* and *Geodia cydonium* showed a nearly symmetrical distribution almost coinciding with the DNA distribution. In this property, the genome of the Demospongiae seems to be very different from those of vertebrates, ranging from fishes to mammals and birds, since the latter are characterized by an asymmetry in the distribution of genes, these features being much more pronounced in warm-blooded vertebrates.

An unexpected result was, however, found when we localized homologous genes shared by the two sponges on the shallow gradient. Tables 3.1 and 3.2 show that there are three pairs of homologous genes in the two sponges: those encoding tetraspanin-CD63R, BHP1 protein and polyubiquitin (the two genes cPKC are not homologous). The sequences of these supposedly orthologous genes extracted from GenBank were aligned with BLAST 2 Sequences (available at www.ncbi.nlm.nih.gov/BLAST/): the two tetraspanin-CD63R genes and the two polyubiquitin genes showed good alignments. Fig. 3.21 shows the localization of these three gene pairs on the *Suberites domuncula* and *Geodia cydonium* shallow gradients, respectively. Contrary to all expectations, the genes BHP1 protein and polyubiquitin are localized on the two fractions in the GC-rich region for *Suberites domuncula*. In contrast, these two genes in *Geodia cydonium* are localized in the GC-poor region of the shallow gradient. Similarly, the tetraspanin-CD63R gene is localized in the GC-poor region of the gradient for *Suberites domuncula* and in the GC-rich region for *Geodia cydonium*.

a)



b)

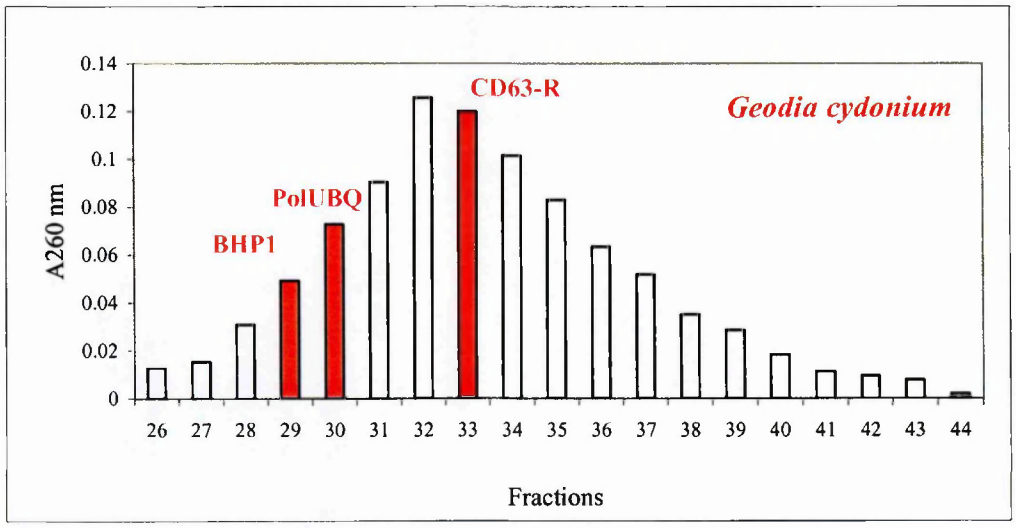


Fig. 3.21 Comparison of localization of the three supposedly orthologous genes (BHP1, PolUBQ and CD63-R) on a) *Suberites domuncula* and b) *Geodia cydonium* shallow gradient.

To understand what happened in the gene distribution, we analyzed the correlations between GC_3 levels of the coding sequences of *Suberites domuncula* and *Geodia cydonium* that had been used in the PCR experiments, and the GC levels of the DNA fractions in which genes were localized. The scatterplots of Fig. 3.22 showed that the slopes of the lines are negative and the correlation coefficients are extremely low.

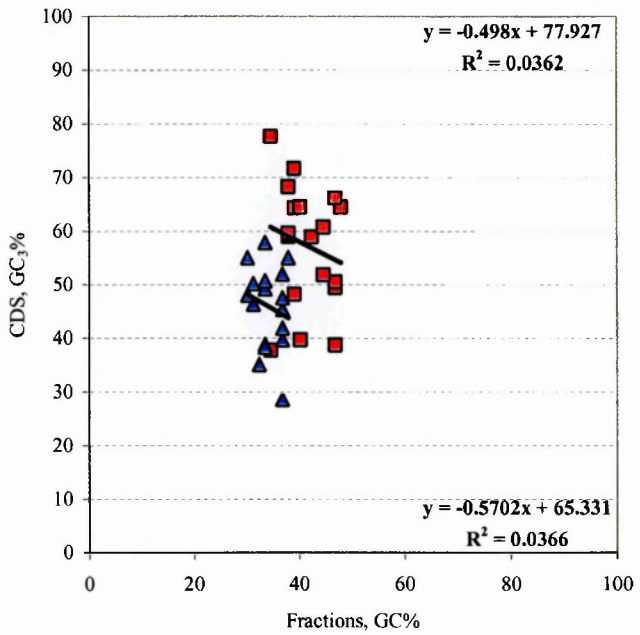


Fig. 3.22 Correlations of GC_3 levels of coding sequences (CDS) versus the GC% of *Suberites domuncula* (blue triangles) and *Geodia cydonium* (red squares) shallow gradient fractions in which the genes are localized.

These results are very unusual because they suggest that there are no correlations between the GC% of the shallow gradient fractions and the GC_3 levels of *Suberites domuncula* and

Geodia cydonium coding sequences. In other words, in these sponges the GC_3-rich genes do not appear to be preferentially located in GC-rich region of DNA, and the GC_3-poor genes do not appear to be preferentially in GC-poor regions.

Since these results may seem surprising, it is relevant to recall what it is known about these types of correlations at this point.

In vertebrate genomes, linear relationships exist between the levels of GC (the molar fraction of guanine + cytosine) or GC_3 (the GC levels of third codon positions) of the coding sequences and the GC levels of the isochores embedding them (Bernardi et al., 1985). Moreover, a correlation exists between GC_3 and GC of coding sequences, which was found to be essentially the same for genes from a number of genomes ranging from bacterial to human (Bernardi and Bernardi, 1985). This was the first suggestion of a general linear relationship between GC_3 and GC_{1+2} (the GC levels of first + second codon positions). In addition, points from different compositional compartments (isochores) of compositionally heterogeneous genomes, such as the genomes of warm-blooded vertebrates, fall on the line of the intergenomic correlations of homogeneous genomes, such as bacterial genomes, showing that the same correlation exists not only intergenomically, but also intragenomically. Further work (Bernardi and Bernardi, 1986) showed that: 1) GC_1, GC_2 and GC_3 values (GC are values pooled from individual prokaryotic and eukaryotic genomes or genome compartments) are positively correlated with the GC levels of the corresponding genomes, a result also reported by Muto and Osawa (1987) for a small sample of bacterial genomes; 2) the slopes of the compositional correlations between individual codon positions and coding sequences were very similar for all classes of

organisms; 3) the frequencies of amino acids change with increasing GC of coding sequences, a point originally made by Sueoka (1961) for bacteria and also reported by Jukes and Bhushan (1986) for bacteria and mitochondria. Further investigations showed that the same correlation holds between GC_3 and GC_{1+2} for human genes (Aïssani et al., 1991; D'Onofrio et al., 1991) and for genes from cold-blooded vertebrates, lower eukaryotes, viruses and bacteria (Bernardi and Bernardi, 1991). Finally, investigations by D'Onofrio and Bernardi (1992) led to the definition of a universal correlation among codon positions both inter- and intra-genomically. The universal correlation was re-analysed on a vastly larger sample of coding sequences and revealed that, in the high GC range of the GC_3 versus GC_1 correlation, there are differences between prokaryotes and eukaryotes. Fig. 3.23 shows the orthogonal regression lines of GC_3 versus GC_1 and GC_2, for prokaryotes, and eukaryotes.

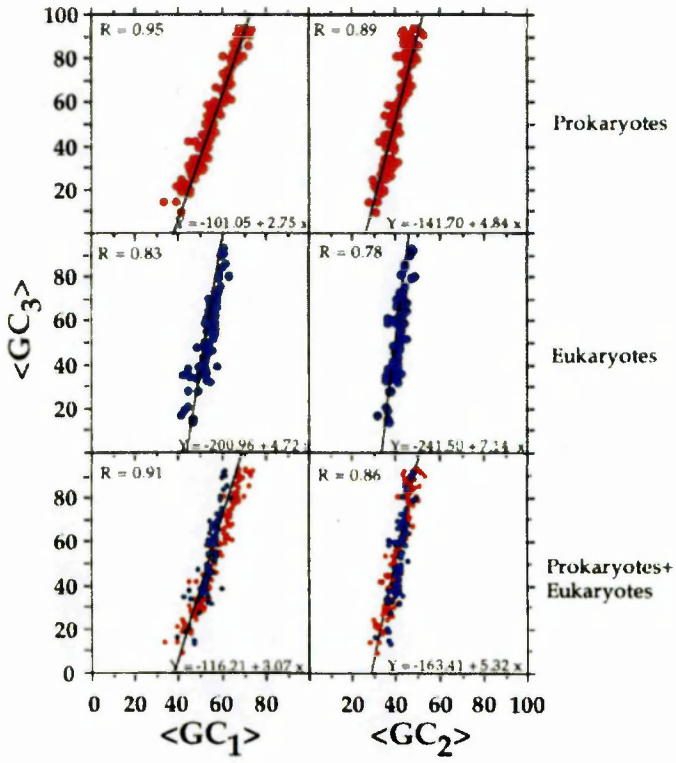


Fig. 3.23 Intergenomic compositional correlations. GC_3 values of genes averaged by genome or genome compartments (in the case of heterogeneous genomes) are plotted against the corresponding GC_1 and GC_2 values. Plots for prokaryotes (red dots), eukaryotes (blue dots) and prokaryotes + eukaryotes are shown, along with the equations of orthogonal regression lines and correlation coefficients (from D'Onofrio et al., 1999).

High correlation coefficients were found in GC_3 versus GC_2 plots for both prokaryotes and eukaryotes. The slopes and intercepts of the orthogonal regressions were slightly higher in eukaryotes compared to prokaryotes, but a standard test (Jolicoeur, 1990) showed that the differences were not significant. The correlations between GC_3 and GC_1 also showed high

coefficients for all prokaryotes and eukaryotes, and the slopes were different for the two groups. Fig. 3.23 also shows the correlation obtained when prokaryotes and eukaryotes are pooled together. Clearly, on a first approximation, a universal correlation still exists between GC_1 and both GC_2 and GC_3. In fact, the equation of the regression line of GC_3 versus GC_{1+2} is not significantly different from that previously published using a small number of genes (D'Onofrio and Bernardi, 1992).

It should be considered that in genes, second position of codons are largely constrained by the amino acids they encode, whereas third positions reflect constraints in base composition. The scatterplot of the frequencies of GC base pairs in the second (GC_2) and third (GC_3) positions of genes from a given genome defines a correlation that is well conserved from prokaryotes to eukaryotes (D'Onofrio et al., 1999). In all species, represented by a large set of experimentally sequenced genes, analyzed to date, the axis is far away from the diagonal ($GC_2 = GC_3$). This conservation was apparently violated in the recently sequenced and annotated rice genome (Yu et al., 2002), which showed many genes aligning along the expected axis, but also many extending along the diagonal. Such behaviour would simply indicate contamination of the data set by intergenic or other noncoding DNA (Cruvellier et al., 2003). Furthermore, 50.6% of genes reported for rice had no orthologs in *Arabidopsis thaliana*. Almost all the genes clustering along the diagonal (Fig. 3.24) were in fact annotated as predicted or putative, whereas the large majority of the experimentally determined genes lined up along the axis that is expected for coding sequences.

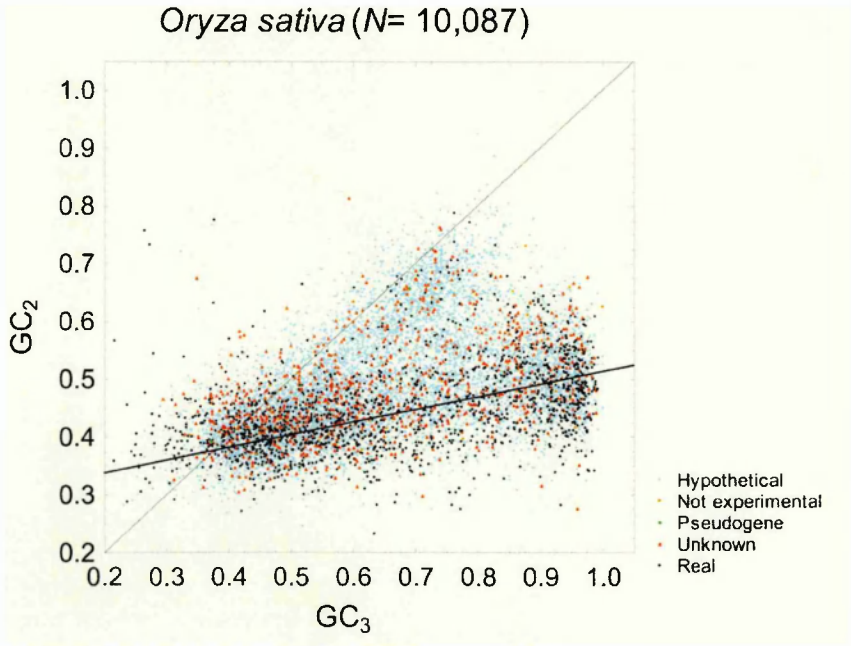


Fig. 3.24 Scatterplot of GC_2 versus GC_3 levels in predicted and experimentally identified rice genes. The diagonal ($GC_2 = GC_3$) is indicated. Complete coding sequences from *Oryza sativa* were extracted from GenBank (release 129; retrieved 31 May 2002) using ACNUC software. Redundancies were removed on the basis of protein alignments using as a cutoff 90% identity for an overlap of 90%. The resulting gene set (N = 10.087) was partitioned into five classes according to the annotations (real genes, not experimental, unknown, pseudogenes and hypothetical) in the informative fields product, gene name, evidence and note, using a script written in Perl (from Cruvellier et al., 2003).

Many, if not most, of the points appearing along the main diagonal in the figure are likely to represent rice sequences that are not translated into proteins. This may have led to

considerably overestimating the proportion of coding sequences that lack orthologs in Arabidopsis. Simple GC_2 versus GC_3 scatterplots can, therefore, serve as a quick check to identify computationally predicted or expressed sequence tag-based genes that are unlikely to code for proteins.

On this basis, complete coding sequences were taken from start codon (ATG) to stop codon and we tested the correlations of GC_1 and GC_2 of *Suberites domuncula* and *Geodia cydonium* coding sequences available in GenBank versus GC_3 (Figs. 3.25 a-b, 3.26 a-b, respectively). The orthogonal regression lines that characterize them are shown, together with the main diagonal of slope 1 (GC_1 = GC3, GC2 = GC_3) as a comparison.

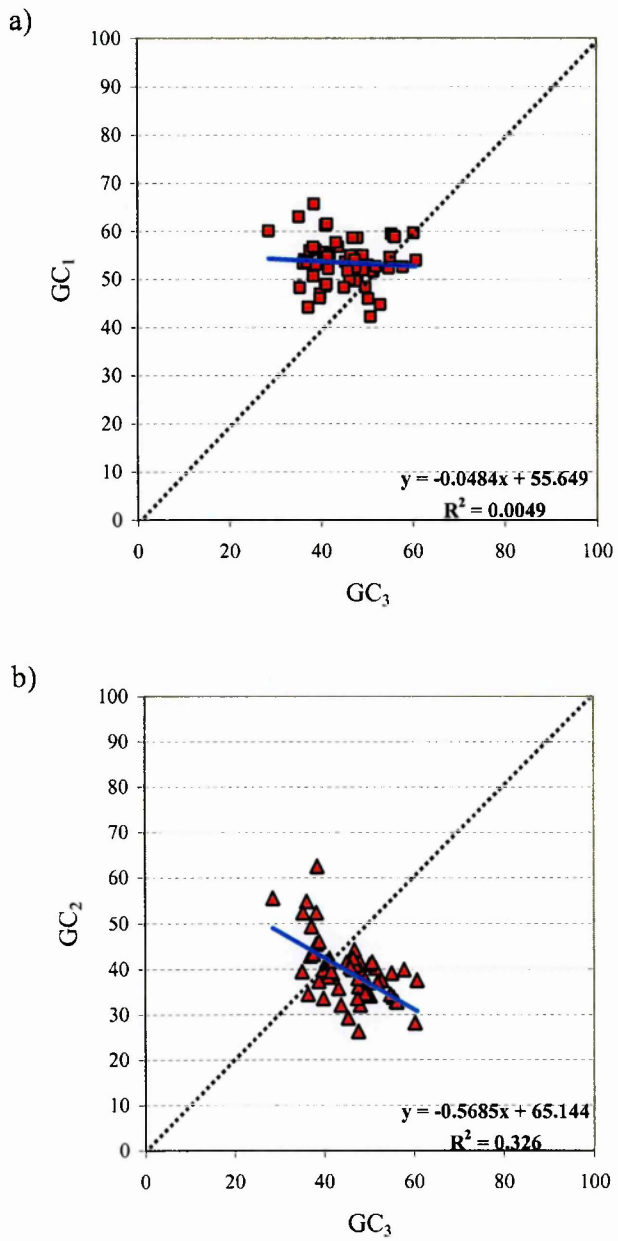


Fig. 3.25 Scatterplot of a) GC_1 versus GC_3 and b) GC_2 versus GC_3 levels of *Suberites domuncula* coding sequences available in GenBank. The main diagonal is also shown.

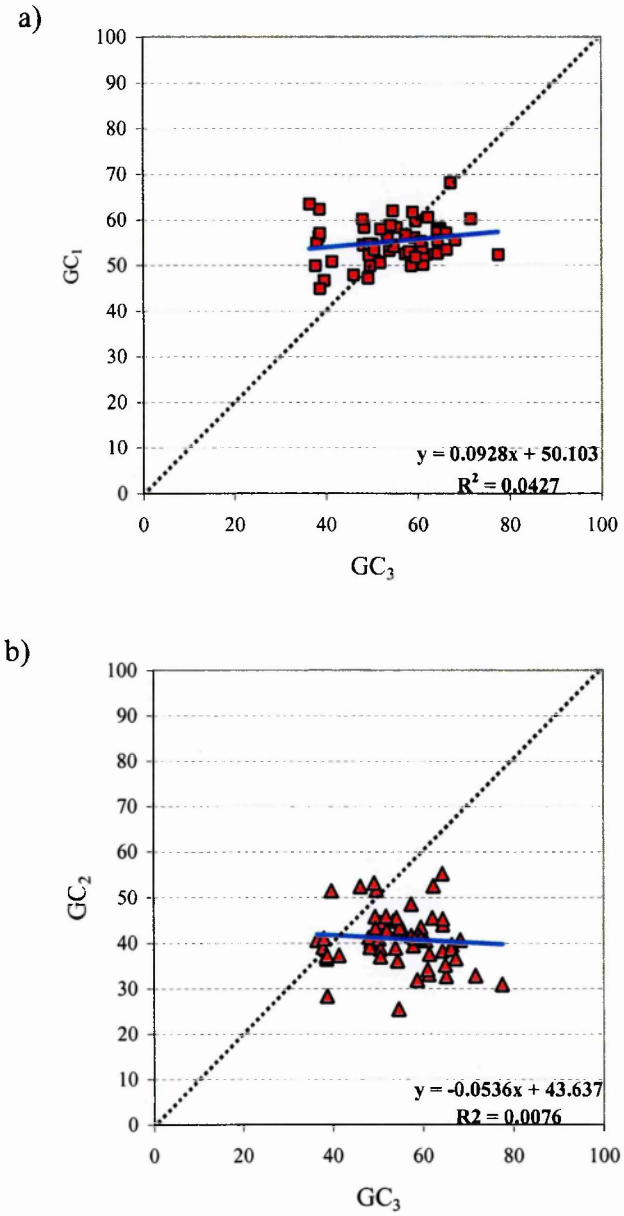


Fig. 3.26 Scatterplot of a) GC_1 versus GC_3 and b) GC_2 versus GC_3 levels of *Geodia cydonium* coding sequences available in GenBank. The main diagonal is also shown.

The correlation coefficient is significant only for the correlation of GC_2 versus GC_3 levels for gene sequences of *Suberites domuncula*, and in this case the correlation seem to be negative. These scatterplots indicate that the universal correlations are not respected in these two sponges and these data go against what it is known in literature. In particular not only we didn't find the universal positive correlations that are well conserved from prokaryotes to eukaryotes (D'Onofrio et al., 1999) but also we are not in the case of the rice genome (Cruvellier et al., 2003) in which this conservation was apparently violated due to contamination of the data set by intergenic or other noncoding DNA.

In Figs. 3.27a-b and 3.28 a-b the same correlations reported in Figs. 3.25a-b and 3.26a-b were reported considering only the genes localized experimentally on the shallow gradient fractions.

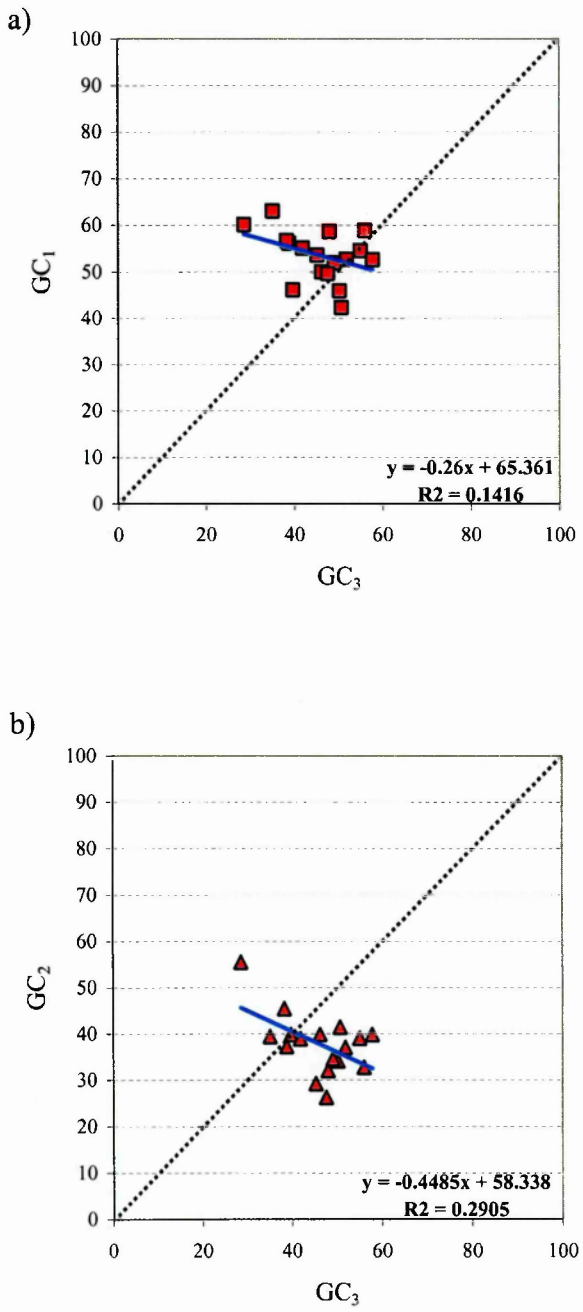


Fig. 3.27 Scatterplot of a) GC_1 versus GC_3 and b) GC_2 versus GC_3 levels of *Suberites domuncula* coding sequences experimentally localized on shallow gradient fractions. The main diagonal is also shown.

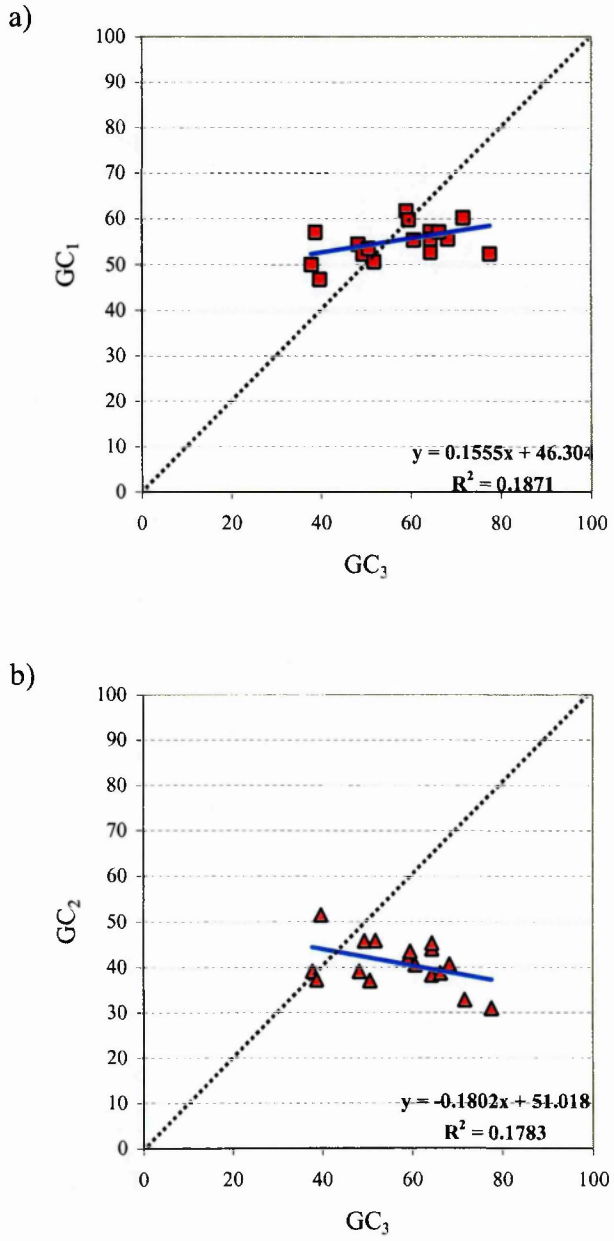


Fig. 3.28 Scatterplot of a) GC_1 versus GC_3 and b) GC_2 versus GC_3 levels of *Geodia cydonium* coding sequences experimentally localized on shallow gradient fractions. The main diagonal is also shown.

As it is possible to see from these scatterplots, the negative correlations found for *Suberites domuncula* is less strong because the points with high GC_2 values didn't localize on shallow gradient fractions; for the others correlations the situation didn't change in a significant way.

For a comparison we can also consider the correlations of coding sequences for human and *Escherichia coli* (Fig. 3.29).

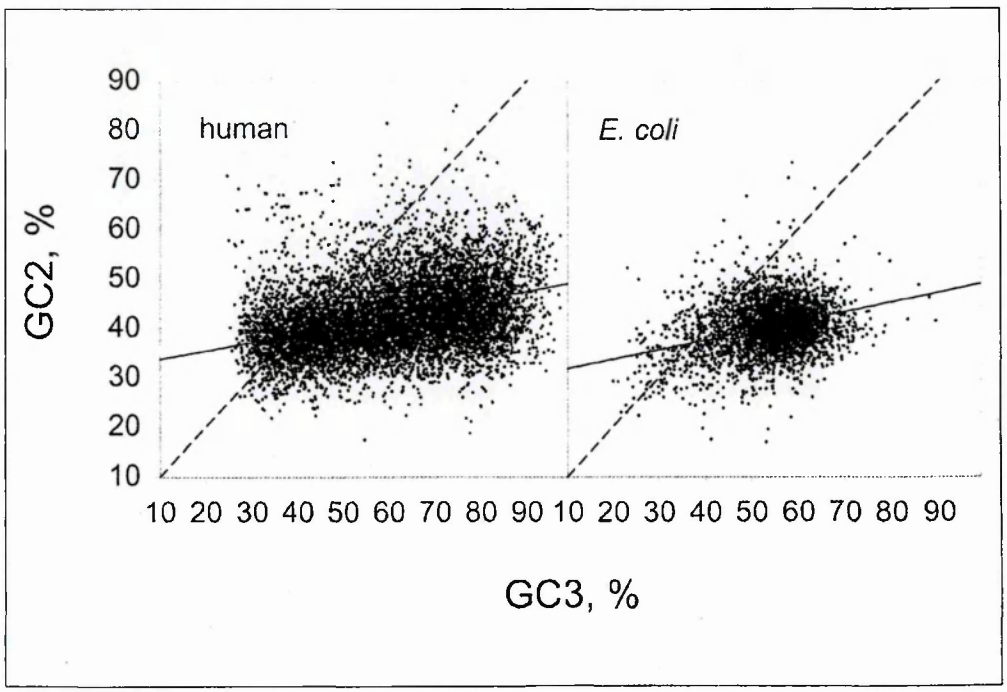


Fig. 3.29 Scatterplots of GC_2 versus GC_3 for non-redundant, representative collections of coding sequences for human (left, 10,128 sequences) and *E. coli* (right, 4,286 sequences). In each scatterplot, the main diagonal and orthogonal regression line are shown. (From Cruvellier et al. 2003).

For human and *Escherichia coli* all the points are along the orthogonal regression line making a very dense cloud and only a small number of points is formed by outliers. Comparing our results with these last correlations, it is possible to observe that in the case of the sponges only a small number of the genes is in the cloud. Considering that there are not a lot of sequences, there are a great number of outliers.

Because of these unusual compositional properties, we tried to understand what happened with the sponge genes. We decided to examine in detail the sequences in GenBank. We analysed the amino acid composition of these genes. In particular we tested the percent of each amino acid because it is known that there are some amino acids that are rare in the usual proteins (for example the aromatic amino acids). From this analysis result only some proteins that have a content of tryptophan, or methionine different from protein usual content.

An analysis at protein levels was done by BLASTX (available at www.ncbi.nlm.nih.gov.). It should be stressed that there were a little number of gene sequences in GenBank of others sponges with which *Suberites domuncula* and *Geodia cydonium* sequences could be aligned. Furthermore, the only significant alignments that we found had a low percentage of identity. As example, was reported the protein tyrosine kinase of *Geodia cydonium* that had 33% of similarity with the protein tyrosine kinase of *Ephydatia fluviatilis*, another sponge that belongs to the class of Demospongiae. Low values of identity (of about 30-40%) were also found with homologous proteins in others organisms, for example with *Drosophila melanogaster*, *Danio rerio*, *Caenorabditis elegans*, *Xenopus laevis* and *Homo sapiens*, that especially due to the phylogenetic distance.

At this point we don't known which type of sponge sequences are those in GenBank. After these analyses it is possible to conclude that *Suberites domuncula* and *Geodia cydonium* coding sequences available in GenBank have problems but it is difficult to understand of which type because there are not enough terms of comparison. It is possible to hypothesize that for some sequences there were problems of frame shift that can be the cause of the reversal of correlations found. On the other hand, we can hypothesize that, concerning the sponge genes, the strange correlations found is because we are in the case of predicted genes.

3.4 Identification of associated organisms

Sponges are probably an extreme example of "infested" organisms because, unlike most other invertebrates, there are no sterile areas in a sponge (Pomponi and Willoughby, 1994). The upper surface area of the sponge (the cortex) is particularly exposed to the contamination. They have two distinct layers, the outer ectosome and the inner endosome. It is in the endosome that some sponges also harbour vast numbers of others organisms (Webb and Maas, 2002). Sponges provide an ideal habitat for microorganisms. Marine sponges frequently contain a complex mixture of bacteria (both symbiotic and incidental), fungi, unicellular algae and cyanobacteria (also both symbiotic and incidental). Significant progress has been made in the documentation of sponge-associated microorganisms and their possible function as endosymbionts.

3.4.1 Bacteria

A brief introduction on the possible type of association among the sponges and their associated organisms will precede the results obtained from this experimental work.

Sponge-bacteria interactions are probably among the oldest host-bacteria interactions known, dating back more than 500 million years (Wilkinson et al., 1984). Several recent studies have revealed that permanent associations exist between certain host sponges and specific micro-organisms, their interactions remaining largely, however, unknown (Preston et al., 1996; Schumann-Kindel 1997; Althoff et al., 1998; Friedrich et al., 1999; Schmidt et al., 2000). Sponges are thought to live in a symbiotic relationship (Simpson, 1984) with unicellular organisms such as prokaryotes, bacteria (Vacelet, 1970) and primarily

cyanobacteria (Vacelet, 1971), eukaryotes, zooxanthellae (yellow symbiotic dinomastogotes) (Sarà and Liaci, 1964) or zoochlorellae (green symbiotic algae) (Gilbert and Allen, 1973). These organisms occur both extracellularly and intracellularly (Wilkinson, 1978).

Virtually all sponges contain endosymbiotic micro-organisms, and these symbionts often contribute considerably to the total sponge biomass (Wilkinson, 1978; Brantley et al., 1995). Before summarising the different type of organisms that have been isolated from *Suberites domuncula* and *Geodia cydonium*, it is necessary to give a few definitions. All micro-organisms found in association with the sponge host will be termed "associated organisms" (Osinga et al., 2001). These can be microbes that are coincidentally present in the sponge, microbes that grow in the mesohyl and microbes that permanently live inside the sponge cells. In addition, it is possible to use the term "symbionts" for those micro-organisms that are always found in association with the same host species. The sponge symbiont relationship can be classified as obligatory mutualism (i.e. the symbionts play an essential role in the metabolism of their host), facultatively mutualism (they have a beneficial effect on their host, but the host will survive without the symbiont) or commensalisms (they are present without providing obvious beneficial effects to their host). In all cases, it is assumed that the sponge host provides a sheltered habitat for their symbionts. A further distinction is made between "epibionts" (micro-organisms living on the sponge surface) and "endosymbionts" (micro-organisms that either live in the sponge mesohyl or inside the sponge cells). A logical question to ask is "why do sponges tolerate micro-organisms inside their body?" The most obvious answer might be that the micro-

organisms provide a source of food or other useful metabolic products to their host. It has been suggested that growth of these useful micro-organisms may be under the control of the sponge host (Muller el al., 1981). This growth of beneficial micro-organisms is termed "gardening" or "farming" and may occur frequently among sponges.

In addition to a transient seawater population serving as a food source, sponge harbor large amounts of bacteria in their tissues that can amount to 40% of their biomass (Vacelet, 1975). Furthermore, sponges may also succumb to microbial and fungal infections which result in the disintegration of the sponge fibers/tissue and ultimately lead to sponge death (Lauckner, 1980; Vacelet et al., 1994).

A very powerful method extensively used to identify symbiotic organisms, especially from those living in a marine ecosystem (Giovanni 1991), is based on PCR amplification of 16S rRNA using universal prokaryotic-specific primers for bacteria 27F-1385R (see Materials and Methods): a fragment of about 1400 bp was amplified. PCR amplification, cloning and subsequent sequencing were performed as described in "Materials and Methods".

Possible correlations between the bacterial population which lives associated with *Suberites domuncula* and *Geodia cydonium* and that of their surrounding water column were investigated. The seawater surrounding the two sponge (15-20 metres in depth) was collected and filtered through a Millipore 0.22 μ filter. These filters were placed on LB (Luria-Bertani medium) agar in ASW plate at 20°C. In this case two bacterial species were isolated from *Suberites domuncula* (Table 3.3, SdB3 and SdB4) and only one from *Geodia*

cydonium (Table 3.4, GcB3). Database searches using the BLASTN program revealed their highest similarity of these clones with the bacterial sequences in GenBank.

The *Suberites domuncula* and *Geodia cydonium* cell suspensions obtained from the dissociated tissue were centrifuged at low speed (600 x g) and both supernatants were plated on LB agar in ASW (artificial seawater) and incubated at 20°C to allow the marine bacteria growth, since these two sponges were collected at this temperature of water column. Five colonies, identifiable from their different colours on the growth plates, were obtained from *Suberites domuncula* (Table 3.5 SdB5, SdB6, SdB7, SdB8, SdB9) and 5 from *Geodia cydonium* (Table 3.6 GcB4, GcB5, GcB6, GcB7, GcB8): they belong to a different bacterial species than those obtained from surrounding water column.

In addition, the bacterial populations of cell suspensions obtained from both dissociated tissue and centrifuged at low speed (600 x g) were analysed. The two genomic DNA were extracted from these two pellets, obtained at 600 x g, and PCR amplification was done. Three clones were isolated from *Suberites domuncula* (SdB10, SdB11, SdB12) and two from *Geodia cydonium* (GcB9 and GcB10). A part of both pellets was also placed on LB agar in ASW plates: three types of colonies were identified for *Suberites domuncula* (SdB13, SdB14 and SdB15) and three for *Geodia cydonium* (GcB11, GcB12 and GcB13).

Table 3.3 Isolated bacterial clones from *Suberites domuncula.*

Bacterial isolate	Source	Highest similarity (%)	Accession number	Buoyant density
SdB3	Water column	Photobacterium sp. KT0248 95%	AF235127	——
SdB4	Water column	Alteromonas sp. MS23 99%	AF237977	1.7063
SdB5	Supernatant cell dissociated	Vibrio natriegens (ATCC 14048T) 97%	X74714	1.7142
SdB6	Supernatant cell dissociated	Marinobacter marinus strain SW-45 99%	AF479689	1.7005
SdB7	Supernatant cell dissociated	Bacillus pumilus 99%	AB098578	1.6978
SdB8	Supernatant cell dissociated	Bacillus sp. VAN35 98%	AF286486	1.7006
SdB9	Supernatant cell dissociated	Bacillus so. OS-5 99%	AJ296095	——
SdB10	DNA from pellet 600xg	Uncultured gamma proteobacterium HOC27 94%	AF384207	——
SdB11	DNA from pellet 600xg	Pseudoalteromonas sp. RE10F/5 94%	AF118019	——
SdB12	DNA from pellet 600xg	Unidentified gamma proteobacterium 94%	AB013824	——
SdB13	Pellet plated	Bacillus hwajinpoejnsis 99%	AJ296095	1.7001
SdB14	Pellet plated	Bacillus decolorationis 97%	AJ315075	1.7002
SdB15	Pellet plated	Bacillus sp. LMG 21002 99%	AJ316308	1.7006

Table 3.4 Isolated bacterial clones from *Geodia cydonium.*

Bacterial isolate	Source	Highest similarity (%)	Accession number	Buoyant Density
GcB3	Water column	North sea bacterium H120 99%	AF069667	——
GcB4	Supernatant cell dissociated	Pseudoalteromonas sp. 93%	AF530129	1.6931
GcB5	Supernatant cell dissociated	Alpha proteobacterium MBIC3368 98%	AF218241	1.7105
GcB6	Supernatant cell dissociated	Bacterium str. 47083 99%	AF227837	1.6988
GcB7	Supernatant cell dissociated	Bacillus hwajinpoensis 99%	AF541966	1.701
GcB8	Supernatant cell dissociated	Alpha proteobacterium MBIC3368 98%	AF218241	1.6999
GcB9	DNA from pellet 600xg	Uncultured gamma proteobacterium HOC2 97%	AB054136	——
GcB10	DNA from pellet 600xg	Uncultured gamma proteobacterium HOC27 94%	AB054161	——
GcB11	Pellet plated	Alpha proteobacterium MBIC3368 99%	AF218241	1.7135
GcB12	Pellet plated	Vibrio sp. OS53 99%	AB038028	1.7037
GcB13	Pellet plated	Vibrio sp. QY101 99%	AY174869	1.7038

All these clones were subjected to phylogenetic analysis. In total, 13 independent sequence profiles were obtained from *Suberites domuncula* and 11 from *Geodia cydonium*. The sequence results indicate that a high diversity of bacterial phylotypes was present within the two sponges. In particular for *Suberites domuncula* 7 clones clustered within the γ-subdivision of the *Proteobacteria* and 6 clones within *Bacillus* (Fig. 3.30).

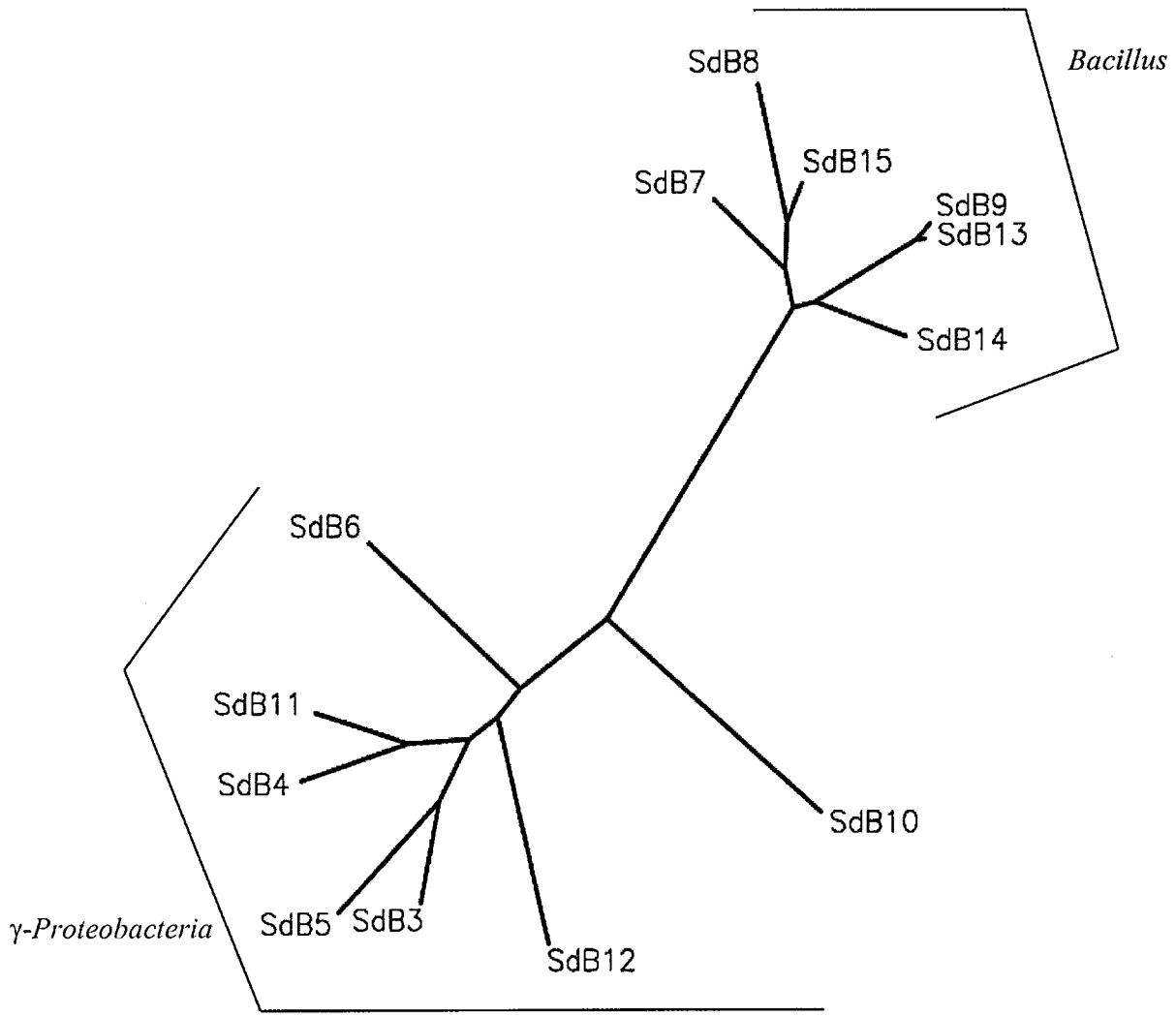


Fig. 3.30 Phylogenetic tree for bacterial clones extracted from *Suberites domuncula*.

For *Geodia cydonium* 3 clones clustered within the α-subdivision of the *Proteobacteria*, 2 within *Bacillus* and 6 within the γ-subdivision of the *Proteobacteria* (Fig. 3.31).

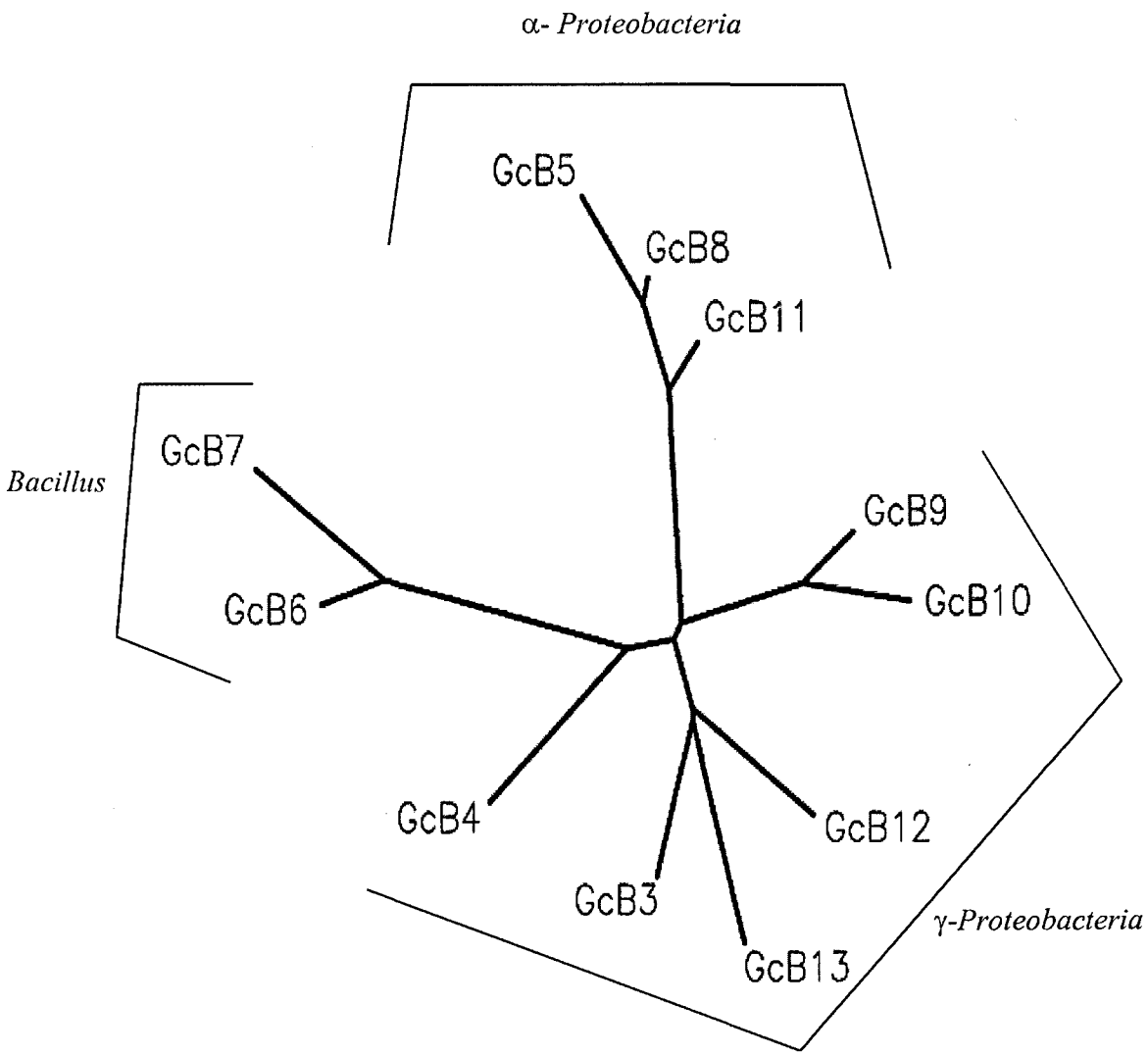


Fig. 3.31 Phylogenetic tree for bacterial clones extracted from *Geodia cydonium*.

In total from the two sponges were isolated 13 of the clones clustered within the γ-subdivision of the *Proteobacteria*, 3 within the α-subdivision of the *Proteobacteria* and 8 within *Bacillus* (see phlylogenetic tree Fig. 3.32).

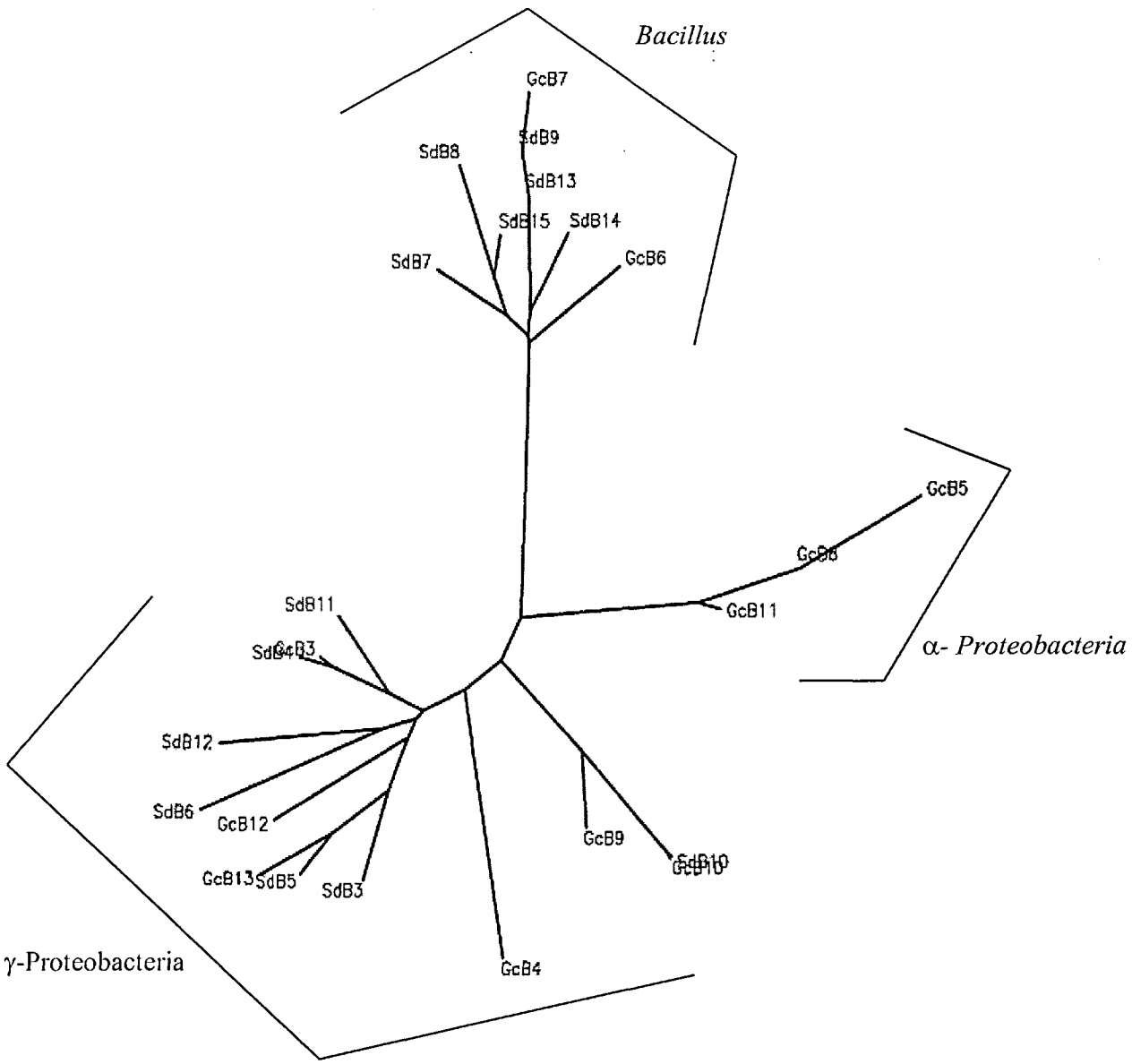


Fig. 3.32 Phylogenetic tree for bacterial clones extracted from *Suberites domuncula* and *Geodia cydonium*.

Webster et al. in 2001 reported similar data. Molecular techniques were employed to document the microbial diversity associated with the marine sponge *Rhopaloeides odorabile.* The community structure was extremely diverse with representatives of the *Actinibacteria*, low GC gram-positive bacteria, the β- and γ- subdivisions of the *Proteobacteria*, *Cytophaga/Flaviobacterium*, green sulphur bacteria, green nonsulphur bacteria, planctomycetes, and other sequence types with no known close relatives.

Firstly, these results strongly suggests that *Proteobacterium sp. Kt0248* and *Alteromonas sp. MS23* (SdB3 and SdB4, respectively) which lives in the *Suberites domuncula* sorrounding water column, are not being utilised as a food source and have not a specific association with *Suberites domuncula*, because they were not found in *Suberites domuncula* (see the other bacterial clones isolated). It is possible to make the same comment on *North sea bacterium H120* (GcB3) in regard to *Geodia cydonium.*

For the analysis of the other bacterial clones it is important to consider the different cellular composition among supernatant and pellet after centrifugation at 600x g. In particular the pellet seems to be enriched for the most part with big (granular) cells, whereas in the supernatant stay small cells. This can explain the different bacteria found when the supernatant and the pellet have been analyzed. On this basis it is possible to suppose that there are some bacteria that prefer living in association with big cells (10 μm) and others that prefer living in association with small cells (2-5 μm). It is possible to suppose that the different sponge cellular populations produce various secondary metabolites that could select between the different bacteria or vice versa. Furthermore, in the supernatant it should be possible to find also the bacteria that live in the intercellular space: they are released

after the tissue dissociation. On the basis of the different sponge symbiont relationship the bacteria SdB5, SdB6, SdB7, SdB8, SdB9, SdB13, SdB14, SdB15 from *Suberites domuncula* and GcB4, GcB5, GcB6, GcB7, GcB8, GcB11, GcB12 and GcB13 from *Geodia cydonium* could be considered extracellular associated organisms or epibionts.
On the contrary, only the bacteria SdB10, SdB11, SdB12, GcB9 and GcB10 could be considered intracellular associated organisms or "endosymbionts" for *Suberites domuncula* and *Geodia cydonium* respectively, because they are released after the cellular lysis that occurs to DNA extraction. These bacteria should be good candidates to be possible obligate symbionts for the two sponges in analysis.

Concerning the bacteria isolated from supernatant of cells dissociated and from pellets after centrifugation at 600 x g plated on LB agar in ASW, their extracted DNA were analyzed by CsCl analytical ultracentrifugation: the buoyant densities of each bacterium are reported in the Table 3.3 for *Suberites domuncula* and in Table 3.4 for *Geodia cydonium*. As it results, all the bacterial ultracentrifugation profiles are under the range of the heterogeneity of *Suberites domuncula* and *Geodia cydonium*. That can explains the diffusion that was observed before in the Fig. 3.6 and in the Fig. 3.11. Under the analytical profiles of both sponges are in hiding the profile of at least 8 bacteria. For these reasons the *Suberites domuncula* and *Geodia cydonium* analytical profiles show 1) a large diffusion, 2) a basiline not on the zero and 3) a tail on the right part.

3.4.2 Archaea

Archaea, one of the three major domains of extant life, are thought to comprise predominantly microorganisms that inhabit extreme environments, inhospitable to most Eucarya and Bacteria. They comprise cultivated members that span a fairly limited range o phenotypes, represented by extreme halophiles, sulfur-metabolizing thermophiles, thermophilic sulfate-reducers and methanogens (DeLong et al., 1992). In the marine environment, archaeal habitats are generally limited to shallow or deep-sea anaerobic sediments (free-living and endosymbiotic methanogens), hot springs or deep-sea hydrothermal vents (methanogens, sulfate reducers, and extreme thermophiles), and highly saline land-locked seas (halophiles).

However, molecular phylogenetic surveys of native microbial assemblages are beginning to indicate that the evolutionary and physiological diversity of Archaea is far greater than previously supposed. Preston et al. in 1996 reported the discovery and preliminary characterization of a marine archeon (*Cenarchaeum symbiosum* gen. no., sp. nov.) that inhabits the tissues of temperate water sponge. The association was specific, with a single crenarchaeal phylotype inhabiting a single sponge host species. This partnership represents the first described symbiosis involving Crenarchaeota. The symbiotic archaeon grows well at temperatures of 10°C, over 60°C below the growth temperature optimum of any cultivated species of Crenarchaeota. Archaea have been generally characterized as microorganisms that inhabit relatively circumscribed niches, largely high-temperature anaerobic environments. In contrast, data from molecular phylogenetic surveys, suggest that some crenarchaeotes have diversified considerably and are found in a wide variety of

lifestyles and habitats. *Cenarchaeum symbiosum* is a symbiotic archaeon closely related to other nonthermophilic crenarchaeotes that inhabit diverse marine and terrestrial environments.

Margot et al. in 2002 described the association between filamentous Archaea and three Mediterranean species of sponges from the family Axinellidae (Porifera: Demospongiae). *Axinella damicornis*, *A. verrucosa* and *Axinella* sp. harbour a high concentration of filamentous Archaea in the collagen that surrounds the siliceous spicules that form their skeleton. Molecular studies have revealed that the filamentous Archaea from the three *Axinella* are closely related and are species specific, with a single phylotype inhabiting each sponge species. They are closely related to *C. symbiosum*, the archaeon found in a sponge from the same genus, *A. mexicana*, although this sponge harbours two phylotypes of the archaeon and they seem to be unicellular (Preston et al., 1996; Schleper et al., 1998). Several attempts have been made to cultivate these Archaea, with no success, suggesting that they may have metabolic needs perhaps only provided by their host sponges.

PCR amplifications with Archaea-specific primers for 16S rDNA (Ar4F/1119aR see Materials and Methods) were done on partially purified *Suberites domuncula* and *Geodia cydonium* genomic DNA. A PCR product of about 1100 bp was obtained only on *Geodia cydonium* DNA. This *Geodia cydonium* PCR product was cloned and 18 clones were sequenced: 11 of these isolated clones resulted closely related to *Uncultured marine archaeal group 1 crenarchaeote clone ST-3k4A* (Accession number AJ347774; similarity of 97%, see phlylogenetic tree Fig. 3.33) and 7 to *Uncultured marine archaeal group 1*

crenarchaeote clone ST-12k16A (Accession number AJ347776; similarity of 97%, see phlylogenetic tree Fig. 3.34), two different strains of single species.

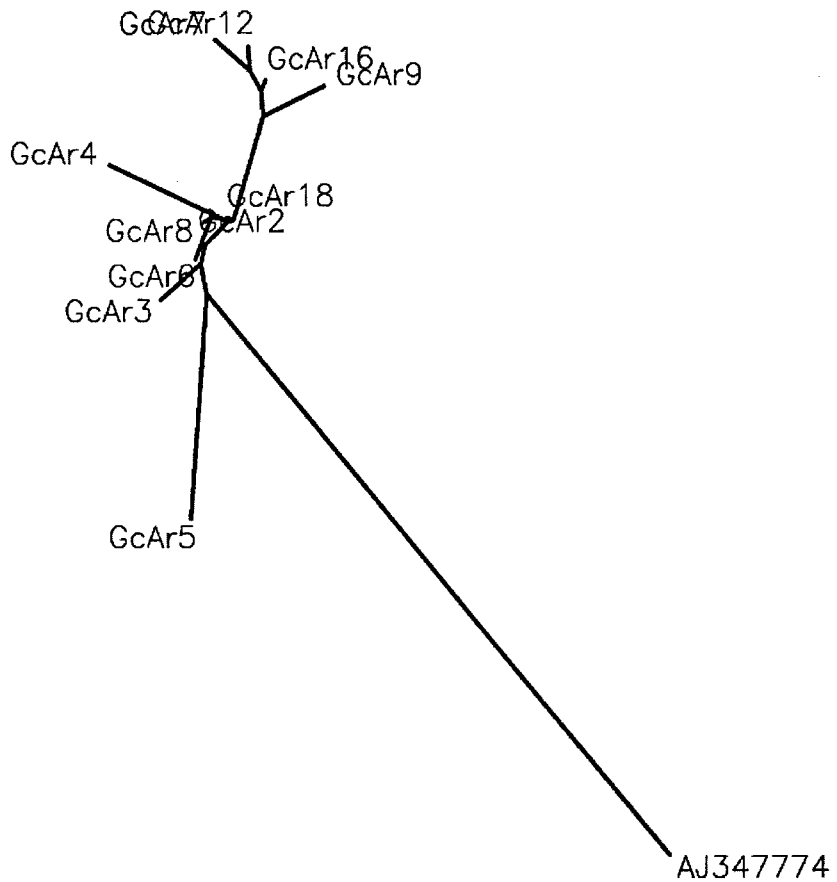


Fig. 3.33 Phylogenetic tree in which are reported 11 of archaea isolated clones closely related to *Uncultured marine archaeal group 1 crenarchaeote clone ST-3k4A* (Accession number AJ347774).

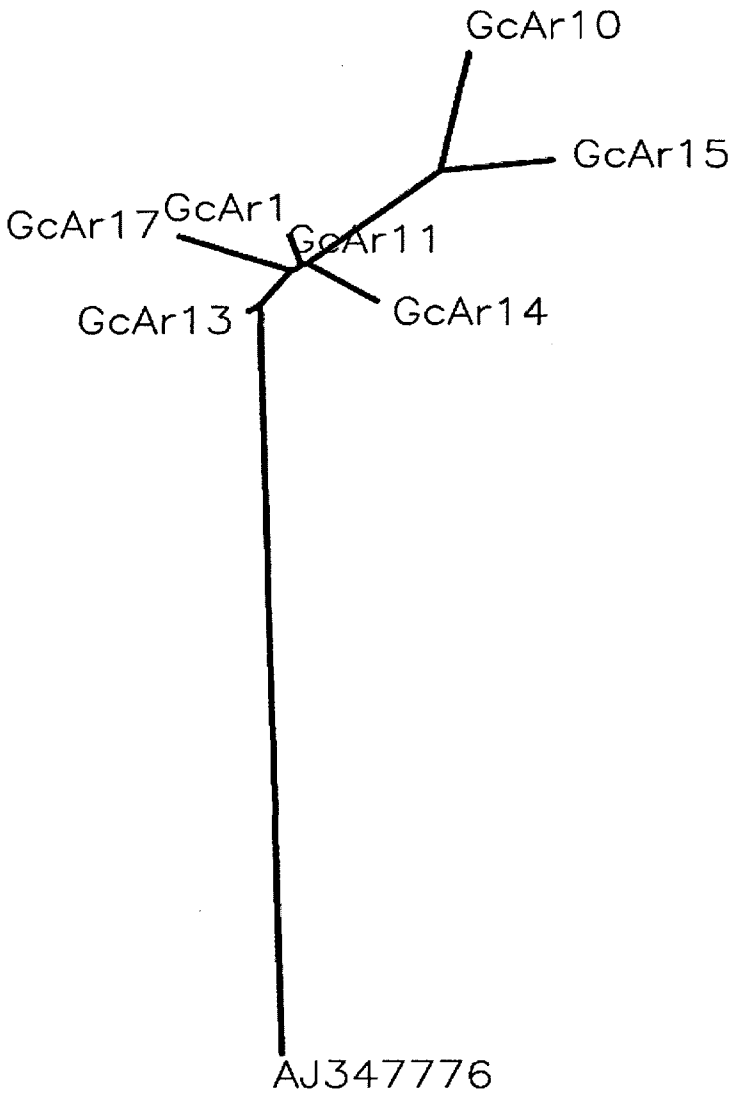


Fig. 3.34 Phylogenetic tree in which are reported 7 of archaea isolated clones closely related to *Uncultured marine archaeal group 1 crenarchaeote clone ST-12k16A* (Accession number AJ347776).

Fig. 3.35 shows the total phylogenetic analysis between all the archaea clones isolated.

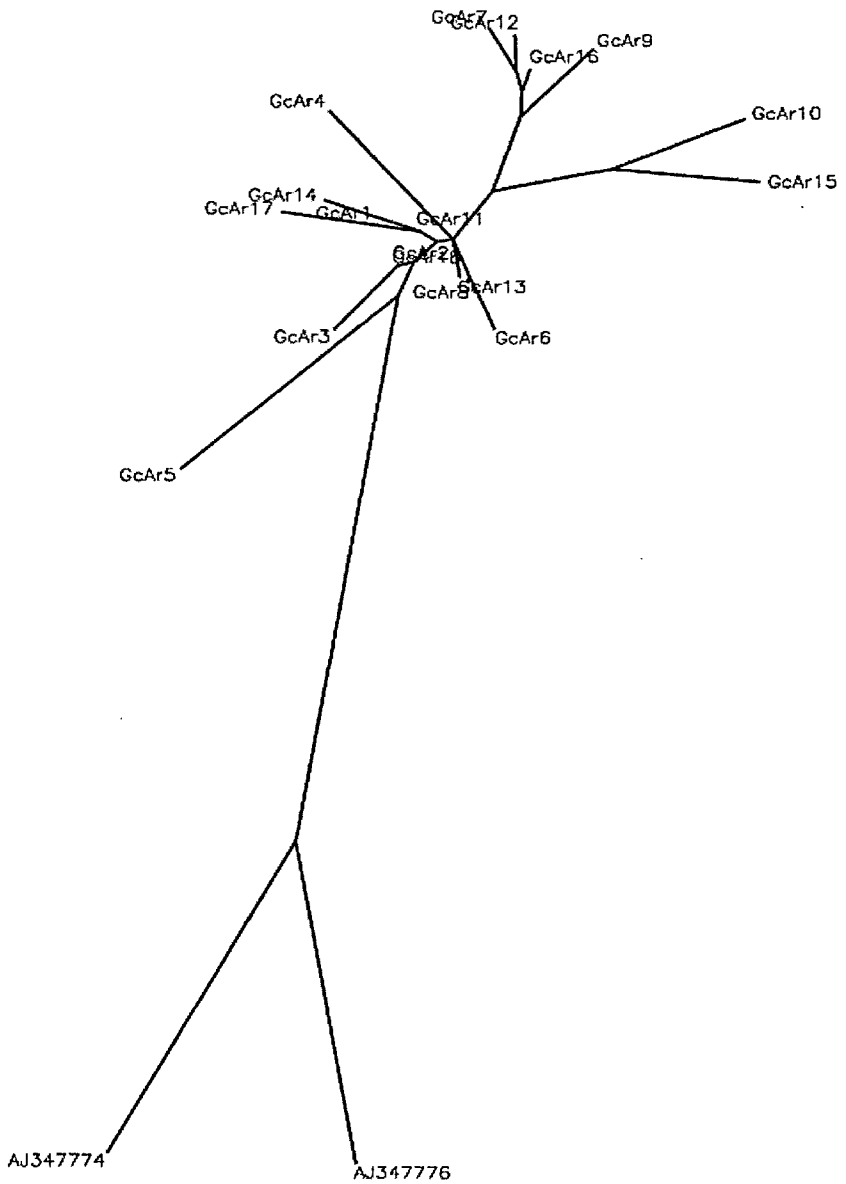


Fig. 3.35 Total phylogenetic tree between all the archaea clones isolated from *Geodia cydonium*.

After a phylogenetic analysis done also in relationship with *Cenarchaeum symbiosum* found by Preston it was possible to state that the clones isolated in this study are not correlated.

The marine "group 1" crenarchaeotes is a newly found group of non-cultivable Archaea that are significant components of marine picoplankton assemblages (DeLong, 1992; DeLong et al., 1999). Several attempts have been made to cultivate these Archaea with no success suggesting that they may have metabolic needs perhaps only provided by their host sponges. The results of this study suggest a novel example of a species-specific symbiosis between *Geodia cydonium* and Archaea in the sea of Naples. It is important to keep in mind that the growth temperature of *Geodia cydonium* in its natural habitat ranges from 10°C to 20°C, and these sponge (and its crenarchaeal symbionts) have remained healthy for months when maintained in laboratory aquaria of our Institute at about 15-20°C. This observation provides strong evidence that the marine crenarchaeotes, whose closest cultivated relatives are all thermophilic or hyperthermophilic, can thrive at low temperatures. Available phylogenetic and ecological data suggest that ancestral variants of hyperthermophilic crenarchaeotes, perhaps originally inhabiting marine hydrothermal systems, became well-adapted for growth in surrounding cold seawater. This colder environment may have been gradually exploited, initially by mesophilic crenarchaeal genetic variants, whose descendants eventually adapted to even lower temperatures of contemporary seas (Preston el al., 1996). Subsequently, mesophilic or psycrophilic crenarchaeotes apparently radiated into many diverse habitats, becoming widespread in marine plankton (Fuhrmann et al., 1992; DeLong et al., 1994), entering into symbiotic associations with metazoa, and

eventually invaded terrestrial environments (Ueda et al., 1995). In analogy to other marine prokaryotic species, nonthermophilic marine Crenarhaeota occupy a wide variety of habitats, ranging from planktonic to symbiotic niches.

3.4.3 Eukaryotes

Suberites domuncula and *Geodia cydonium* genomic DNA extracted from whole tissue was used to amplify and clone the rDNA fragment between two universal eukaryotic primers (ITS3 and D2), corresponding to a highly variable region of the molecule (Fig. 3.36).

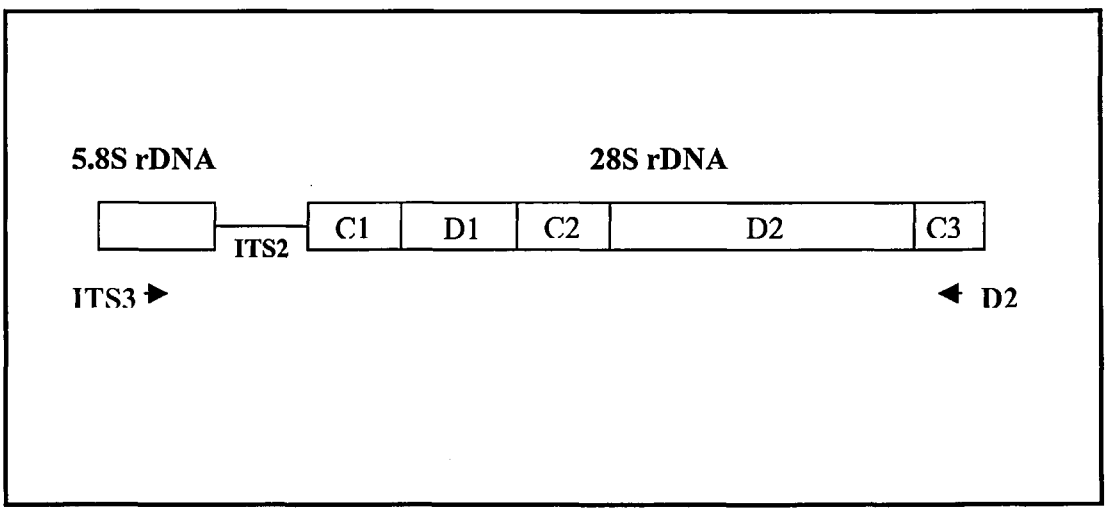


Fig. 3.36 Structure of rDNA and localization of two universal eukaryotic primers, ITS3 and D2.

Cloning and sequencing of the ITS3-D2 fragment should allow to verify whether eukaryotic DNA other than that of the sponge is present in the preparation. A PCR product of about 1200 bp was obtained. At present, 20 clones have been sequenced: all clone sequences result identical to the sequence of *Suberites domuncula*. Probably that means

Eukaryotes are not present in *Suberites domuncula.* Similar analysis done on *Geodia cydonium* revealed the presence of two eukaryotic clones, called GcEu1 and GcEu2 respectively. BLAST search showed that GcEu1 displays the highest similarity to *Chattonella subsalsa* (Eukaryota; Stramenopiles; Raphidiphyceae; Chattonella) with approximatively 92 % similarity, instead GcEu2 has the highest similarity to *Chlorarachnion CCMP621* (Eukaryota; Cercozoa; Chlorarachniophyceae; Chlorarachnion) with approximatively 89 % similarity. Concerning *Chattonella subsalsa* is an heterokont alga and may be involved in harmful algal blooms. Indeed, concerning *Chlorarachnion CCMP621* belongs to the Chlorarachniophytes that are green amoeboflagellate algae that are primarily distinguished by the presence of a plastid of secondary endosymbiotic origin (Keeling 2001). Primary plastids (those of plants, green algae, red algae and glaucocystophytes) arose through the endosymbiotic uptake of a cyanobacterium by a eukaryote, but the ancestor of chlorarachniophytes acquired its plastid by swallowing a photosynthetic eukaryote and, rather than simply digesting it as food source, retaining the alga to perform photosynthesis. Now the algal endosymbiont is severely reduced and is completely integrated with its amoeboflagellate host such that the two are regarded as a single organism (McFadden and Gilson 1995). The origins of both the host and the endosymbiont components of chlorarachniophytes have proved to be quite puzzling, since both are unusual and extremely highly adapted to their endosymbiotic association. Before secondary endosymbiotic plastid origin was understood, it was thought that *Chlorarachnion* was likely a relative of heterokont algae (Keeling, 2001); however, plastid pigmentation eventually suggested that the endosymbiont was some kind of green alga.

This has recently been confirmed by molecular phylogeny (Ishida et al., 1997), but still no strong evidence from either pigmentation or molecular data could demonstrate conclusively what kind of green alga it was. Indeed, when *Chlorarachnion* was first discovered, the presence of a plastid naturally tempted investigators to suggest that the whole cell was related to other algal groups. However green algal origin of chlorarachniophyte plastids was recognized.

There is in the literature some evidence of sponge/algae association. For example *Ephydatia fluviatilis* is a freshwater sponge that harbours algae. In particular, this sponge shows variations of its green pigmentation according to light intensity and seasonality (Corallini and Gaino, 2001). Sponge pigmentation is related to the presence of endocellular zoochlorellae that are restricted to the mesohyl cells (mainly archeocytes) of the outermost layers of the sponge. Symbionts reside in individual membrane-limited cytoplasmic vacuoles; commonly there is only a single element per cells. The ultrastructural organisation of the algae within these cells testifies to their progressive digestion by the host. Occasionally, intact zoochlorellae appear between sponge cell pseudopodia before becoming included into vacuoles.

Bugni et al. in 2002 reported the data about the association of the red macro alga *Ceratodictyon spongiosum* and its sponge symbiont *Haliclona cymaeformis*.

Chapter 4

- Conclusions -

The first part of this research project was devoted to analyse the GC level heterogeneity of the DNA in genomes of the two sponges *Suberites domuncula* and *Geodia cydonium* that belong to the class of Demospongiae.

Because in the literature there were some evidences of organisms that live in symbiosis with these two sponges which cannot be easily separated from the sponge tissue, the first step was the purification of sponge DNA. Firstly we obtained two CsCl analytical ultracentrifugation profiles for both sponges in analysis (Figs. 3.3-3.5) that showed three peaks, suggesting an extreme heterogeneity of both DNA or the presence of associated organisms. It should be consider that the only data present in the literature about the heterogeneity of the sponge DNA were reported from Bartmann et al. in 1997 concerning *Geodia cydonium* DNA. The authors showed an analytical profile having an extreme heterogeneity never observed before for any organism. Applying different protocols with particular precaution, it was possible to obtain partial DNA purification for both sponges. In particular, it was possible for us, for the first time, to obtain CsCl analytical ultracentrifugation profiles for *Suberites domucula* (Fig. 3.6) and *Geodia cydonium* (Fig. 3.11) DNA that showed one peak that is due to the sponge DNA, characterized by different values of buoyant density (ρ = 1.6987 g/cm^3 for *Suberites domuncula*; ρ = 1.7031 g/cm^3 for *Geodia cydonium*). The other two peaks, due certainly to the presence of associated organisms, were eliminated although not completely. However they are not visible in CsCl analytical ultracentrifugation profiles. We calculated from the buoyant

density of the CsCl analytical profiles, using the equation of Schildkraut et al. (1962), the GC% of both DNA, corresponding to 39.6 for *Suberites domuncula* DNA and 43.9 for *Geodia cydonium* DNA.

The second aim of this experimental work was to assess the gene distribution in the genome of these two sponges. The base composition heterogeneity of sponge DNA allows this DNA to be fractionated by CsCl density gradient centrifugation, using the "shallow gradient" technique. As results we obtained shallow gradient fractionations which showed 19 fractions for *Geodia cydonium* DNA (Fig. 3.17) and 25 fractions for *Suberites domuncula* DNA (Fig. 3.18).

The next step was the analysis of the gene sequences in GenBank to choose the genes to analyse. PCR amplification with specific primers was used to localize genes of interest in GC-poor or GC-rich genome DNA fractions. PCR conditions were optimized for 17 genes for *Suberites domuncula* and 18 for *Geodia cydonium*. Each of these genes was localized on the shallow gradient fractions (see Fig. 3.20a-b). After this type of the analysis we have a series of strange results. The localization of the analysed coding sequences from both *Suberites domuncula* and *Geodia cydonium* showed a nearly symmetrical distribution almost coinciding with the DNA distribution. In this property, the genome of the Demospongiae seems to be very different from those of vertebrates, ranging from fishes to mammals and birds, since the latter are characterized by an asymmetry in the distribution of genes, these features being much more pronounced in warm-blooded vertebrates.

An unexpected result was, however, found when homologous genes shared by the two sponges on the shallow gradient were localized. Tables 3.1 and 3.2 show that there are

three pairs of homologous genes in the two sponges: those encoding tetraspanin-CD63R, BHP1 protein and polyubiquitin. Fig. 3.21 shows the localization of these three gene pairs on the *Suberites domuncula* and *Geodia cydonium* shallow gradients, respectively. Contrary to all expectations, the genes BHP1 protein and polyubiquitin are localized on the two fractions in the GC-rich region for *Suberites domuncula.* In contrast, these two genes in *Geodia cydonium* are localized in the GC-poor region of the shallow gradient. Similarly, the tetraspanin-CD63R gene is localized in the GC-poor region of the gradient for *Suberites domuncula* and in the GC-rich region for *Geodia cydonium.*

To understand what happened in the gene distribution, we analyzed the correlations between GC_3 levels of the coding sequences of *Suberites domuncula* and *Geodia cydonium* that had been used in the PCR experiments, and the GC levels of the DNA fractions in which genes were localized (Fig. 3.22): the slopes of the lines are negative and the correlation coefficients are extremely low. These data went against the universal correlation existing of GC_3 versus GC_1 and GC_2 (D'Onofrio el al., 1999). In fact, high correlation coefficients were found in GC_3 versus GC_2 plots for both prokaryotes and eukaryotes. The correlations between GC_3 and GC_1 also showed high coefficients for all prokaryotes and eukaryotes. These correlations resulted well conserved from prokaryotes to eukaryotes (Fig. 3.23). It needs to be considered that this conservation was apparently violated only in the rice genome (Fig. 3.24), which showed many genes aligning along the expected axis, but also many extending along the diagonal, indicating contamination of the data set by intergenic or other noncoding DNA (Cruvellier et al., 2003).

On this basis, we tested the correlations of GC_1 and GC_2 of *Suberites domuncula* and *Geodia cydonium* coding sequences available in GenBank versus GC_3 (Figs. 3.25 a-b,

3.26 a-b). The orthogonal regression lines that characterize them are shown, together with the main diagonal of slope 1 (GC_1 = GC3, GC2 = GC_3) as a comparison. The correlation coefficient is significant only for the correlation of GC_2 versus GC_3 levels for gene sequences of *Suberites domuncula*, and in this case the correlation seem to be negative. These scatterplots indicate that the universal correlations are not respected in these two sponges and these data go against what it is known in literature. In particular not only we didn't find the universal positive correlations that are well conserved from prokaryotes to eukaryotes (D'Onofrio et al., 1999) but also we are not in the case of the rice genome (Cruvellier et al., 2003) in which this conservation was apparently violated due to contamination of the data set by intergenic or other noncoding DNA. Moreover, it should be stressed that that we are in an unusual case in which for the first time the range of the GC_2 is about the same of that of GC_3 (with a range of about 30%): usually in all the organisms till now studied GC_2<GC1 and GC_2<< GC_3 (except viruses which show the same degree of constraint at all the three codon position because of the overlapping reading frame). Also considering only the sponge genes localized experimentally (Figs. 3.27a-b, 3.28 a-b), the scatterplots showed that the negative correlations found for *Suberites domuncula* is less strong because the points with high GC_2 values didn't localize on shallow gradient fractions; for the others correlations the situation didn't change in a significant way.

Because of these unusual compositional properties we decided to examine in detail the sequences in GenBank. We analysed the amino acid composition of these genes. In particular we tested the percent of each amino acid and from this analysis result only

some proteins that have a content of tryptophan, or methionine different from protein usual content.

An analysis at protein levels was done by BLASTX It should be stressed that there were a little number of gene sequences in GenBank of others sponges with which *Suberites domuncula* and *Geodia cydonium* sequences could be aligned. Furthermore, the only significant alignments that we found had a low percentage of identity, that especially due to the phylogenetic distance.

At this point we don't known which type of sponge sequences are those in GenBank. After these analyses it is possible to conclude that *Suberites domuncula* and *Geodia cydonium* coding sequences available in GenBank have problems but it is difficult to understand of which type because there are not enough terms of comparison. It is possible to hypothesize that for some sequences there were problems of frame shift that can be the cause of the reversal of correlations found. On the other hand, we can hypothesize that, concerning the sponge genes, the strange correlations found is because we are in the case of predicted genes.

The last part of the study was devoted to the identification of associated organisms, in particular bacteria, Archaea and Algae. The advances in molecular biology have provided new and important diagnostic possibilities, not only for the classification of prokaryotes but also for the determination of phylogenetic relationships among animals. The gene sequences, which most commonly have been used, are 16S rRNA for the analysis of bacteria. The preceding observations, made in species that are markedly different systematically, morphologically, and ecologically, show that the occurrence of intimately associated bacteria is a general phenomenon in sponges and that various

aspects of the association are different to the species studied. One of surprising findings that come out of this study is the discovery of a sponge-specific, yet phylogenetically diverse, microbial community. The phylogenetic signature of the sponge-associated microbial consortium is distinctly different from that of typical seawater. The molecular taxonomic analysis of sponge-associated bacteria from *Suberites domuncula* and *Geodia cydonium* indicates that there is a diverse assemblage of bacteria residing within these sponges; however, none of these previously cultured microorganisms were identified in the present study. In particular, 13 bacterial clones were isolated from *Suberites domuncula* and 11 from *Geodia cydonium*: 13 of the clones clustered within the γ-subdivision of the *Proteobacteria*, 3 within the α-subdivision of the *Proteobacteria* and 8 within *Bacillus* (see phlylogenetic tree Fig. 3.32). It was possible to hypothesize the different types of relationships that these bacterial clones had with the sponges. Bacteria SdB5, SdB6, SdB7, SdB8, SdB9, SdB13, SdB14, SdB15 from *Suberites domuncula* and GcB4, GcB5, GcB6, GcB7, GcB8, GcB11, GcB12 and GcB13 from *Geodia cydonium* could be considered extracellular associated organisms or epibionts (see Tables 3.3-3.4). Bacteria SdB10, SdB11, SdB12, GcB9 and GcB10 could be considered intracellular associated organisms or "endosymbionts" for *Suberites domuncula* and *Geodia cydonium* respectively and should be good candidates to be possible obligate symbionts. The observed microbial pattern reflects instead an adaptation to the specific conditions of the sponge mesohyl tissue. Environmental factors are responsible for the creation of this ecological niche.

Concerning the Archaea, only in *Geodia cydonium* were isolated. In particular, 11 of these isolated clones resulted closely related to *Uncultured marine archaeal group 1*

crenarchaeote clone ST-3k4A (Fig. 3.32) and 7 to *Uncultured marine archaeal group 1 crenarchaeote clone ST-12k16A* (Fig. 3.33). Several attempts have been made to cultivate these Archaea with no success suggesting that they may have metabolic needs perhaps only provided by their host *Geodia cydonium.*

Lastly, searching for the presence of Eukaryotes we found two algal clones *Chattonella subsalsa*, an heterokont alga involved in harmful algal blooms, and *Chlorarachnion CCMP621*, that is a green amoeboflagellate alga.

References

Aissani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K., Gautier, C., Bernardi G., 1991. The compositional properties of human genes. J. Mol. Evol., 32, 497-503.

Althoff, K., Schütt, C., Steffen, R., Batel, R., Müller, W.E.G., 1998. Evidence for a symbiosis between bacteria of the genus *Rhodobacter* and the marine sponge *Halichondria panicea*: harbor also for putatively toxic bacteria? Mar. Biol. 130, 529-536.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Ac. Res. 25, 3389-3402.

Bartmann-Lindholm, C., Geisert, M., Gungerich, U., Muller, W.E.G., Weinblum, D., 1997. Nuclear DNA fractions with grossly different base ratios in the genome of the marine sponge *Geodia cydonium*. Progr. Colloid. Polym. Sci. 107, 122-126.

Bernardi, G., 1965. Chromatography of nucleic acids on hydroxyapatite. Nature (London), 206, 779-783.

Bernardi, G., 1989. The isochore organization of the human genome. Annu. Rev. Genet. 23, 637-661.

Bernardi, G., 1990. Le Génome des Vertébrés: organisation, function et evolution. Biofutur 94, 43-46.

Bernardi, G., 1993a. The human genome organization and its evolutionary history - A review. Gene 135, 57-66.

Bernardi, G., 1993b. Genome organization and species formation in vertebrates. J. Mol. Evol. 37, 331-337.

Bernardi, G., 1993c. The vertebrate genome: isochores and evolution. Mol. Biol. Evol. 10, 186-204.

Bernardi, G., 1995. The human genome: organization and evolutionary history. Annu. Rev. Genet. 29, 445-476.

Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3-17.

Bernardi, G., Bernardi, G., 1986. Compositional constraints and genome evolution. J. Mol. Evol. 24, 1-11.

Bernardi, G., Bernardi, G., 1990a. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 265-281.

Bernardi, G., Bernardi, G., 1990b. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 282-293.

Bernardi, G., Bernardi, G., 1991. Compositional properties of nuclear genes from cold-blooded vertebrates. J. Mol. Evol. 33, 57-67.

Bernardi, G., Mouchiroud, D., Gautier, C., Bernardi, G., 1988. Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. 28, 7-18.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., Rodier, F., 1985. The mosaic genome of warm-blooded vertebrates. Science 228, 953-958.

Blumbach, B., Diehl-Seifert, B., Seak, J., Steffen R., Müller, I.M., Müller, W.E.G., 1999. Cloning and expression of the new receptor belonging to the immunoglobin superfamily from the marine sponge *Geodia cydonium.* Immunogenetics 49, 751-763.

Blumbach, B., Pancer, Z., Diehl-Seifert, B., Steffen, R., Münkner, J., Müller, I.M., Müller, W.E.G., 1998. The putative sponge aggregation receptor. Isolation and characterization of a molecule composed of scavenger receptor cysteine-rich domains and short consensus repeats. Journal of Cell Science 111, 2635-2644.

Böger, H., 1988. Versuch über das phylogenetische System der Porifera. Meyniana 40, 143-154.

Böhm, M., Hentschel, U., Friedrich, A.B., Fieseler, L., Steffen, R., Gamulin, V., Müller, I.M., Müller, W.E.G., 2001. Molecular response of the sponge *Suberites domuncula* to bacterial infection. Mar. Biol. 139, 1037-1045.

Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., Le Parco, Y., 2001. Sponge paraphyly ant the origin of Metazoa. J. Evol. Biol. 14, 171-179.

Brantley, S.E., Molinski, T.F., Preston, C.M., DeLong, E.F., 1995. Brominated acetylenic fatty acids from *Xestospongia* sp., a marine sponge-bacteria association. Tetrahedron 51, 7667-7672.

Bugni, T.S., Concepcion, G.P., Mangalindan, G.C., Harper, M.K., James, R.D., Ireland, C.M., 2002. P-sulfooxyphenylpyruvic acid from the red macro alga *Ceratodictyon spongiosum* and its sponge symbiont *Haliclona cymaeformis*. Phytochemistry 60, 361-363.

Cacciò, S., Perani, P., Saccone, S., Kadi, F., Bernardi, G., 1994. Single-copy sequence homology among the GC-richest isochores of the genomes from warm-blooded vertebrates. J. Mol. Evol. 39, 331-339.

Canfield, D.E., Teske, A., 1998. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. Nature 382, 127-132.

Carels, N., Bernardi, G., 2000. Two classes of gene in plants. Genetics 154, 1819-1825.

Cavalier-Smith, T., 1991. Intron phylogeny: a new hypothesis.Trends Gen. 7, 145-148.

Cavalier-Smith, T., Alsopp, M.T.E.P., Chao, E.E., Boury-Esnault, N., Vacelet, J. 1996. Sponge phylogeny, animal monophyly, and the origin of the nervous system:18S rRNA evidence. J. Zool. 74, 2031-2045.

Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K.G., Adoutte, A., 1991. An analysis of the origin of metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triploblasts. European Molecular Biolgy Organization Journal 10, 499-503.

Clay, O., Cacciò, S., Zoubak, S., Mouchiroud, D., Bernardi, G., 1996. Human coding and non-coding DNA: compositional correlations. Mol. Phylogenet. Evol. 5, 2-12.

Clay, O., Douday, C.J., Carels, N., Hughes, S., Bucciarelli, G., Bernardi, G., 2003a. Using analytical ultracentrifugation to study compositional variation in vertebrate genomes. European Biophysics Journal 32(5), 418-26.

Clay, O., Douday, C.J., Carels, N., Hughes, S., Bucciarelli, G., Bernardi, G., 2003b. Points to watch in CsCl work. European Biophysics Journal (*in press*).

Clay, O., Carels, N., Douady, C., Macaya, G., Bernardi, G., 2001. Compositional heterogeneity within and among isochors in mammalian genomes. I. CsCl and sequence analyses. Gene 276, 15-24.

Cocito, C. 1969. The Action of Virginiamycin on Nucleic Acid and Protein Synthesis in *Bacillus subtilis* Infected with Bacteriophage 2C. J. Gen. Microbiol. 57, 195-206.

Corallini, C., Gaino, E., 2001. Peculiar digestion patterns of sponge-associated zoochlorellae in the cadisfly *Ceraclea fulva*. Tissue and Cells 33(4), 402-407.

Christen, R., Ratto, A., Barion, A., Perasso, R., Grell, K.G., Adoutte, A., 1991. An analysis of the origin of metezoan, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triphoblasts. Eur. Mol. Biol. Org. J. 10, 499-503.

Cruveiller S., Jabbari, Clay O., Bernardi G., 2003. Compositional features of eukaryotic genomes for checking predicted genes. Briefings in Bioinformatics 4(1) 43-52.

Cruveiller S., Jabbari, Clay O., Bernardi G., 2003. Incorrectly predicted genes in rice? Science (in press)

Cuny, G., Soriano, P., Macaya, G. and Bernardi, G., 1981. The major components of the mouse and human genomes. I. Preparation, basic properties and compositional heterogeneity. Eur. J. Biochem. 115, 227-233.

Custodio, M.R., Prokic, I., Steffen, R., Koziol, C., Borojevic, R., Brümmer, F., Nickel, M., Müller, W.E.G, 1998. Primmorphs generated from dissociated cells of the sponge *Suberites domuncula*: a model system for studies of cell froliferation and cell death. Mechanism of Ageing and Development 105, 45-59.

DeLong, E.F., 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA 89, 5685-5689.

DeLong, E.F., Taylor, L.T., Marsh, T.L., Preston, C.P., 1999. Visualization and enumeration of marine planktonic Archaea and bacteria by using polyribonucleotide probes and fluorescent in situ hybridisation. Appl. Environ. Microbiol. 65, 5554-5563.

DeLong, E.F., Wu, K.Y., Prezelin, B.B., Jovine, R.V.M., 1994. High abundance of Archaea in Antarctic marine picoplankton. Nature 371, 695-697.

De Sario, A., Geigl, E.M., Bernardi G., 1995. A rapid procedure for the compositional analysis of yeast artificial chromosomes. Nucleic Acid Research 23 (19), 4013-4014.

D'Onofrio, G., Bernardi, G., 1992. A universal compositional correlation among codon positions. Gene 110, 81-88.

D'Onofrio, G., Jabbari, K., Musto, H., Alvarez-Valin, F., Cruveiller, S., Bernardi, G., 1999a. Evolutionary genomics of vertebrates and its implications. Ann. N.Y. Acad. Sci. 870, 81-94.

D'Onofrio, G., Jabbari K., Musto H., Bernardi G., 1999. The correlation of protein hydropathy with the base composition of coding sequences. Gene 238, 3-14.

D'Onofrio, G., Mouchiroud, D., Aïssani, B., Gautier, C., Bernardi, G., 1991. Correlation between the compositional properties of human genes, codon usage and amino acid composition of proteins. J. Mol. Evol. 32, 504-510.

Dubilier, N., Amann, R., Erséus, C., Muyzer, G., Park, S.Y., Giere, O., Canavaugh, C.M., 1999. Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olvius loisae* (Anellida). Mar. Ecol. Prog. Ser. 178, 271-280.

Elyakov, G.B., Kuznetsova, t., Kikhailov, V.V., Maltsev, I.I., Voinov, V.G., Fedoreyev, S.A., 1991. Brominated diphenyl ethers from a marine bacterium associated with the sponge *Dysidea* sp. Experientia 47, 632-633.

Fasman, G.D., 1976. Handbook of Biochemistry and Molecular Biology; 3rd Edition; volume II, CRC Press; Cleveland, Ohio; p. 290.

Faulkner, D.J., 1995. Marine natural products. Nat. Prod. Rep. 13, 259-302.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., Raff, R.A., 1998. Molecular phylogeny of the animal kingdom. Science 239, 748-753.

Filipski, J., Thiery, J.P., Bernardi G., 1973. An analysis of the bovine genome by Cs_2SO_4-Ag+ density gradient centrifugation. J. Mol. Biol. 80, 177-197.

Flowers, A.E., Garson, M.j., Webb, R.I., Dumdei, E.J., Charan, R.D., 1998. Cellular origin of chlorinated diketopiperazines in the dictyoceratid sponge *Dysidea herbacea* (Keller). Cell Tissue Res. 292, 597-607.

Friedrich, A.B., Mekert, H., Fendert, T., Hacker, J., Proksch, P., Hentschel, U., 1999. Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analyzed by fluorescence *in situ* hybridization (FISH). Mar. Biol. 134, 461-470.

Fuhrmann, J.A., McCallum, K., Davis, A.A., 1992. Novel major archaebacterial group from marine plankton. Nature 356, 148-149.

Gast, R.J., Caron, D.A., 1996. Molecular phylogeny of symbiotic dinoflagellates from planktonic foraminifera and radiolaria. Mol. Biol. Evol. 13(9), 1192-1197.

Gatti, S., 2002. High Antarctic carbon and silicon cycling: how much do sponges contribute? Boll. Mus. Ist. Biol. Univ. Genova 66-67 (2000-2001), 76.

Gilbert, J.J., Allen H.L., 1973. Chlorophyll and primary productivity of some green, freshwater sponges. Int. Revue Ges Hydrobiol. 58, 633-658.

Gilbert, W., 1978.Why genes in pieces? Nature 271, 501.

Giovanni, S.J. 1991. The polymerase chain reaction. In: Stackebrandt E, Goodfellow M (Eds) Nucleis acid techniques in bacterial systematics. John Wiley and Sons Ltd, New York, pp. 177-203.

Grigioni, S., Boucher-Rodoni, R., Demarta, A., Tonolla, M., Peduzzi, R., 2000. Phylogenetic characterisation of bacterial symbionts in the accessory nidamental glands of the sepioid *Sepia officinalis* (Cephalopoda: Decapoda). Marine Biology 136, 217-222.

Grigioni, S., Boucher-Rodoni, R., Tonolla, M., Peduzzi, R., 1999. Symbiotic relations between bacteria and cephalopods. Boll. Soc. Ticinese Sci. nat. 97, 51-53.

Hadzi, J., 1963. The evolution of the Metazoa. Pergamon, Oxford.

Hackel, E., 1868. Natürliche Schöpfungsgeschichte. Reimer, Berlin.

Hentschel, U., Hopke, J., Horn, M., friedrich, A.B., Wagner, M., Haecker, J., Moore, B.S., 2002. Molecular evidence for a uniform microbial communityvin sponges from different oceans. Applied and Environmental Microbiology 68, 4431-4440.

Ilan, M., Contini, H., Carmeli, S., Rinkevich, B., 1996. Progress towards cell cultures from a marine sponge that produces bioactive compounds. J. Mar. Biotechnol. 4, 145-149.

Imsiecke, G., Pascheberg, U., Muller, W.E.G., 1993. Preparation and kayotype analysis of mitotic chromosomes of the freshwater sponge *Spongilla lacustris*. Chromosoma 102, 724-727.

Imsiecke, G., Custodio, M., Borojevic, R., Steffen, R., Moustafa, M.A., Muller, W.E.G. 1995. Genome size and chromosomes in marine sponges [*Suberites domuncula, Geodia cydonium*]. Cell Biology International 19(19), 995-1000.

Ishida, K., Cao, Y., Hasegawa, M., Okada, N., Hara, Y., 1997. The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. J. Mol. Evol. 45, 682-687.

Jolicoeur, P., 1990. Bivariate allometry: interval estimation of slopes of the ordinary and standardized normal major axes and structural relationship. J. Theor. Biol. 144, 275-285.

Jukes, T.H., Bushan, V., 1986. Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J. Mol. Evol. 24, 39-44.

Jurgens, G., Glockner, F., Amann, R., Saano A., Montonen, L., Likolammi, M., Munster, U., 2000. Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridisation(1). FEMS Microbiol. Ecol. 34(1), 45-56.

Keeling, P.J., 2001. Foraminifera and Cercozoa are related in actin phylogeny. Two orphans find a home? Mol. Biol. Evol. 18(8), 1551-1557.

Kelly-Borges, M., Pomponi S.A. 1994. Phylogeny and classification of lithistid sponges (Porifera: Demospongiae): a preliminary assessment using ribosomal DNA sequence comparisons. Mol. Mar. Biol. Biotech. 3(2), 87-103.

Kemp, S., Kazmierczak, J., 1994. The role of alkalinity in the evolution of ocean chemistry, organization of living system, and biocalcification processes. Bull. Inst. Ocean Monaco 13, 61-117.

Knoll, A.H., Carroll, S.B., 1999. Early animal evolution: emerging views from comparative biology and geology. Science 284, 21229-2137.

Koziol, C., Borojevic, R., Steffen, R., Müller, W.E.G., 1998. Sponges (Porifera) model system to study the shift from immortal to senescent somatic cells: the telomerase activity in somatic cells. Mechanism of Ageing and Development 100, 107-120.

Krasko, A., Lorenz, B., Batel, R., Schröder, H.C., Müller, I.M., Müller, W.E.G., 2000. Expression of silicatein and collagen genes in the marine sponge Suberites domuncula is controlled by silicate and myotropin. Eur. J. Biochem. 267, 4878-4887.

Krasko, A., Scheffer, U., Koziol, C., Pancer, Z., Batel, R., Badria, F.A., Müller, W.E.G., 1997. Diagnosis of sublethal stress in the marine sponge Geodia cydonium: application of the 70 kDa heat-shock protein and a novel biomarker, tha Rab GDP dissociation inhibitor, as probes. Aquatic toxicology 37, 157-168.

Krasko, A., Schröder, H.C., Hassanein, H.M.A., Batel, R., Müller, I.M., Müller, W.E.G., 1998. Identification and expression of the SOSresponse, *aidB*-like, gene in the marine sponge *Geodia cydonium*: implication for the phylogenetic relationship of metazoan acyl-coA deydrogenases and acyl-coA oxidases. J. Mol. Evol. 47, 343-352.

Krasko, A., Schröder, H.C., Perovic, S., Steffen, R., Kruse, M., Reichert, W., Müller, I.M., Müller, W.E.G., 1999. Ethylene modulates gene expression in cells of the

marine sponge *Suberites domuncula* and reduces the degree of apoptosis. The Journal of Biological Chemistry 29, 31524-31530.

Kruse, M., Gamulin, V., Cetkovic, H., Pancer, Z., Müller, I.M., Müller, W.E.G., 1996. Molecular evolution of the metazoan protein kinase C mutigene family. J. Mol. Evol. 43, 374-383.

Lafay, B., Boury-Esnault N., Vacelet, J., Christen, R., 1992. An analysis of partial 28S ribosomal RNA sequences early radiations of sponges. Biosystem 28, 139-151.

Lake, J.A., 1990. Origin of Metazoa. Proc. Natl. Acad. Sci. U.S.A. 87, 763-766.

Lauchner, G., 1980. Diseas of Porifera. In: Kinne, O., (ed) Diseases of marine animals. Wiley, Chichester.

Li, W.H., Graur, D. 1991. Fundamentals in molecular evolution. Sinnaeur, Sunderland.

Lipps, J.H., Signor, P.W., 1992. Origin and early evolution of Metazoa. Plenum Press, New York.

Macaya, G., Thiery, J.P., Bernardi G., 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J. Mol. Biol. 108, 237-254.

Margot, H., Acebal, C., Toril, E., Amils, R., Fernandez Puentes, J.L., 2002. Consistent association of crenarchaeal Archaea with sponges of the genus *Axinella*. Marine Biology 140, 739-745.

Margulis, L., Schwartz, K.V., 1995. Five kingdoms: an illustrated guide to the phyla of life on earth. Freeman. New York.

Marmur, J., Doty, P., 1959. Heterogeneity in deoxyribonucleic acids: I. Dependence on composition of the configurational stability of deoxyribonucleic acids. Nature 183, 1427-1429.

McFadden, G., Gilson, P.R., 1995. Something borrowed, something green: lateral transfer of chloroplast by secondary endosymbiosis. Trends Evol. Ecol. 10, 12-17.

Medina, M., Collins, A.G., Silberman, J.D., Sogin, M.L., 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. . PNAS 98(17), 9707-9712.

Medlin, L.K., Williams, D.M., Sims, P.A., 1993. The evolution of the diatoms (Bacillariophyta). Eur. J. Phycol. 28, 261-275.

Mehl, D, Reiswig H.M., 1991. The presence of flagellar vanes in choanomeres of Porifera and their possible phylogenetic implications. Z. Zool. Syst. Evol. Forsch. 29, 312-319.

Mehl, D., 1992. Die Entwicklung der Hexactinellida seit dem Mesozoikum; Paläobiologie, Phylogeneie und Entwicklungsökologie. Berl Geowiss Abh (E) 2, 1-164.

Meselson, M., Stahl, F., Vinograd, J., 1957. Equilibrium sedimentation of macromolecules in density gradients. Proc. Natl. Acad. Sci 43, 581-588.

Moscona, A.A., 1968. Cell aggregation: properties of specific cell-ligands and their role in the formation of multicellular systems. Developmental Biology 18, 250-277.

Mouchiroud, D., Bernardi, G., 1993. Compositional properties of coding sequences and mammalian phylogeny. J.Mol.Evol. 37, 109-116.

Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Gautier, C., Bernardi, G., 1991. The distribution of genes in the human genome. Gene 100, 181-187.

Müller, C. I., Blumbach, B., Krasko, A., 2001. Receptor protein-tyrosine phosphatases: origin of domains (catalytic domain, Ig-related domain, fibronectin type III module) based on the sequence of the sponge Geodia cydonium. Gene 262, 221-230.

Müller, W.E.G., 1995. Molecular phylogeny of Metazoa (animals): monophyletic origin. Naturwissenschaften 82, 321-329.

Müller, W.E.G., 1997. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium*. Cell Tissue Res. 289, 383-395.

Müller, W.E.G., 1997. Evolution of Protozoa to Metazoa. Theory Bioscienc. 116, 145-168.

Müller, W.E.G., 1998. Origin of Metazoa: Sponges as Living Fossils. Naturwissenschaften 85, 11-25.

Müller, W.E.G., 1999a. Establishment of a primary cell culture from a sponge: primmorphs from Suberites domuncula. Marine Ecology Progress Series 178, 205-219.

Müller, W.E.G., 1999b. 16S rRNA sequences as diagnostic tools to elucidate potential symbiotic relationship between bacteria and the marine sponge Halicondria panicea: eply to Stackebrandt and Pukall. Marine Biology 134, 213-215.

Müller, W.E.G., 2001. Review: how was metazoan threshold crossed? The hypothetical Urmetaoza. Comparative Biochemistry and Physiology Part A 129, 433-460.

Müller, W.E.G., 2003. Molecular biodiversity. Case study: Porifera (sponges). Naturwissenschaften 90, 103-120.

Müller, W.E.G., Böhm, M., Grebenjuk, V. A., Skorokhod, A., Müller, I.M., Gamulin, V., 2002. Conservation of the positions of metazoan introns from sponges to human. Gene 295, 299-309.

Müller, W.E.G., Müller, I., Kurelec, B., Zahn, R.K., 1976. Species-specific aggregation factor in sponges. Exp. Cell Res. 98, 31-40.

Müller, W.E.G., Müller, I., Pondeljak, V., Kurelec, B., Zahn, R.K., 1978. Species-specific aggregation factor in sponges. Differentiation 10, 45-53.

Müller, W.E.G., Müller, I.M., Gamulin, V., 1994. On the monophyòetic evolution of the Metazoa. Brazil. J. Med. Biol. Res. 27; 2083-2096.

Müller, W.E.G., Zahn, R.K., 1973. Purification and characterization of a specie-specific aggregation factor in sponges. Exp. Cell Res. 80, 95-104.

Müller, W.E.G., Zahn, R.K., Müller I 1981. Cell aggregation of the marine sponge Geodia cydonium. Identification of lectin-producing cells. Eur. J. Cell Biol. 24, 28-35.

Müller, W.E.G., Wiens, M., Batel, R., Steffen, R., Schröder, H.C., Borojevic, R., Custodio, M.R., 1999. Establishment of a primary cell culture from a sponge: primmorphs from *Suberites domuncula*. Mar. Ecol. Prog. Ser. 178, 205-219.

Müller, W.E.G., Wimmer, W., Schatton, W., Böhm, M., Batel, R., Filic, Z., 1999. Initiation of an aquaculture of sponges for the sustainable production of bioactive metabolites in open system: example, *Geodia cydonium*. Mar. Biotechnol. 1, 569-.579.

Murakami, Y., Oshima, Y., Yasumoto, T., 1982. Identification of okadaic acid as a toxic component of a marine dinoflagellate *Prorocenrum lima*. Nihon Sisan Gakkaishi 48, 69-72.

Musto, H., Romero, H., Zavala, A., Bernardi, G., 1999. Compositional correlations in the chicken genome. J. Mol. Evol. 49, 325-329.

Muto, A., Osawa, S., 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Roc. Natl. Acad. Sci. USA 84, 166-169.

Nübel, U., Garcia-Pichel, F., Muyzer, G. 1997. PCR primers to amplify 16S rRNA genes from Cyanobacteria. Applied and Environmental Microbiology 63(8), 3327-3332.

Ohno, S., 1970. Evolution by gene duplication. Springer, Berlin Heidelberg New York.

Osinga, R., Armstrong, E., Burgess, J.G., Hoffmann, F., Reitner, J., Schumann-Kindel, G., 2001. Sponge-microbe associations and their importance for sponge bioprocess enginnering. Hydrobiologia 461, 55-62.

Osinga, R., Tramper, J., Wijffels, R.H., 1999. Cultivation of marine sponges: a (re)view. Mar. Biotechnology 1, 509-532.

Pahler, S., Blumbach, BG., Müller, I.M., Müller, W.E.G., 1998. Putative multiadhesive protein from the marine sponge *Geodia cydonium*: cloning of the cDNA encoding a fibronectin-, an SRCR-, and a complement control protein module. The Journal of Experimental Zoology 282, 332-343.

Pancer, Z., Kruse, M., Müller, I.M., Müller, W.E.G., 1997. On the origin of metazoan adhesion receptors: cloning of integrin α subunit from the sponge *Geodia cydonium*. Mol. Biol. Evol. 14, 391-398.

Pancer, Z., Skorokhod, A., Blumbach, B., Müller, W.E.G., 1998. Multiple Ig-like featuring genes divergent within and among individuals of the marine sponge *Geodia cydonium*. Gene 207, 227-233.

Patthy, L., 1995. Protein evolution by exon-shuffling. Springer Verlag, New York.

Peppard, J.V., Loo, P., Sills, M.A., Munster, D., Pomponi, S.A., Wright, A.E 1996. Characterization of an interleukin 6 cytokine family antagonist protein from a marine sponge, *Callyspongia sp.* J. Biol. Chem. 271(13), 7281-7284.

Perez, T., Garrabou, J., Sartoretto, S., Harmelin, J.G., Francour, P., Vacelet, J., 2000. Mortalitè massive d'invertèbrès marins: un èvènement san precedent en Mediterranée nord-occidentale. CR Acad. Sci. Paris (Sci Vie) 323, 853-865.

Perovic, S., Krasko, A., Prokic, I., Muller, I.M., Muller, W.E.G., 1999. Origin of neuronal-like receptors in Metazoa: cloning of a metabotropic glutamate/GABA-like receptor from the marine sponge *Geodia cydonium*. Cell Tissue Res. 296, 395-404.

Perovic, S., Schröder, H.C., Sudek, S., Grebenjuk, V.A., Batel, R., Stianic, M., Müller, I.M., Müller,W.E.G., 2003. Expression of one sponge *Iroquois* homeobox gene in primmorphs from *Suberites domuncula* during canal formation. Evol. Devel. (*in press*).

Pfeifer, K., Frank, W., Schröder, H.C., Gamulin, V., Rinkevich, B., Batel, R., Muller, I.M., Muller, W.E.G. 1993. Cloning of the polyubiquitin cDNA from the marine sponge *Geodia cydonium* and its preferential expression during reaggreation of cells. J. Cell Science 106, 545-554.

Pomponi, S.A., 1979, Ultrastucture and cytochemistry of the etching area of boring sponges. In: Levi, C., Boury-Esnault, N. (eds) Biologie des spongiaires. Colloq. Int. CNRS 2914, 319-323.

Pomponi, S.A., Willoghby, R., 1994. Sponge cell culture for production of bioactive metabolites. In Van soest, R.W.M., Van Kempen M.G. and Braekman, J.C. (eds) Sponge in Time and Space. A.A. Balkema, Rotterdam: 395-400.

Preston, C.M., Ke Ying Wu, Molinski, T.F. De Long, E.F., 1996. A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov. Proc. Natl. Acad. Sci. USA 93, 6241-6246.

Proksch, P., Edrada, R.A., Ebel, R., 2002. Drugs from the seal: current status and microbiological implications. Appl. Microbiol. Biotechnol. 59, 125-134.

Reiswig, H., 1971. In situ pumping activities of tropical demospongiae. Mar. Biol. 9: 38-50.

Richter, C., Wunsch, M., Rasheed, M., Kötter, I., Badran, M.I., 2001. Endoscopic exploration of Red Sea coral reefs dense populations of cavity-dwelling sponges. Nature 413, 726-730.

Rolfe, R., Meselson, M., 1959. The relative homogeneity of microbial DNA. Proc. Nat. Acad. Sci. USA 45, 1039-1043.

Rodrigo. A.G., Bergquist, P.L., Reeves, R.A., 1994. Are sponges animals? An investigation into the vagaries of phylogenetic interference. Soest RWM v, Kempen TMG V, Braekman JC (eds) Sponges in time and space, Balkema, Rotterdam, pp. 47-54.

Rottmann, M., Schröeder, H.C., Gramzow, M., Renneisen, K., Kurelec, B., Dorn, A., Friese, U., Müller, W.E.G., 1987. Specific phosphorilationof proteinsin pore complex-laminae from the sponge *Geodia cydonium* by the homologous aggregation factor and phorbol ester. Role of protein kinase C in the phosporilation of DNA topoisomerase II. Embo J. 6(13), 3939-3944.

Rowan, R., Powers, D.A., 1991. Molecular genetic identification of symbiotic dinoflagellates (zooxanthellae). Mar. Ecol. Prog. Ser. 71, 65-73.

Rützler, K., 1965. Systematik und Okologieder Poriferen aus Litoral-Schattengebieten der Nordadria. Z. Morphol. Oekol Tiere 55, 1-82.

Rützler, K., Reiger, G., 1973. Sponge burrowing fine structure of Cliona lampa penetrating calcareous substrata. Mar. Biol. 21, 144-162.

Sabeur, G., Macaya, G., Kadi, F., Bernardi, G., 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. J.Mol.Evol. 37, 93-108.

Salomon, C.E., Deerinck, T., Ellisman, M., Faulkner, D.J., 2001. The cellular localization of dercitamide in the Palauan sponge *Oceanapia sagittaria*. Mar. Biol. 139, 313-319.

Sambrook, J., E.F. Fritsch, Maniatis, T. 1989. Molecular cloning: a laboratory manual, Second edition. 3 Volumes. Cold Spring Harbor Laboratory Press, New York.

Sarà, M., Liaci, L., 1964. Symbiotic association between zooxanthellae and two marine sponges of the genus *Cliona*. Nature 203, p. 321.

Sarma, A.S., Daum, T., Müller, W.E.G., 1993. Secondary metabolites from marine sponges. Akademie gemeinnütziger Wissenschaften zu Erfurt, Ullstein-Mosby, Berlin.

Scheffer, U., Krasko, A., Pancer, Z., Müller, W.E.G., 1997. High conservation of the serum response factor within Metazoa: cDNA rom the sponge *Geodia cydonium*. Biological Journal of the Linnean Society 61, 127-137.

Schleper, C., DeLong, E., Preston, C.M., Feldman, R.A., Ke Ying Wu, Swanson, R.V., 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum sumbiosum. Journal of Bactreriology 180(19), 5003-5009.

Schildkraut, C. L., Marmur, J., Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. 4, 430-443.

Schmidt, E.W., Obraztsova, A.Y., Davidson, S.K., Faulkner, D.J., Haygood, M.G., 2000. Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel δ-proteobacterium, "*Candidatus* Entotheonella paluensis". Mar. Biol. 136, 969-977.

Schröder, H.C., Kruse, M., Batel, R., Müller, I.M., Müller, W.E.G., 2000. Cloning and expression of the sponge longevity gene *SDLAGL*. Mechanisms of Development 95, 219-220.

Schuhmann-Kindel, G., Bergbauer, M., Reitner, J., 1997. Bacteria associated with Mediterranean sponges. In: reitner, J., Neuweiler, F., Gunkel, F. (eds) Global and regional controls on biogenic sedimentation. Reef evolution. Research Reports. Goett Arb Geol Paläont 8b2, 125-128.

Schütze, J., Skorokhod, A., Müller, I.M., Müller, W.E.G., 2001. Molecular evolution of the metazoan extracellular matrix: cloning and expression of structural proteins from the Demosponges *Suberites domuncula* and *Geodia cydonium*. J. Mol. Evol. 53, 402-415.

Seack, J., Kruse, M., Müller, W.E.G., 1998. Evolutionary analysis of G-proteins in early metazoan: cloning of α- and β-subunits from the sponge *Geodia cydonium*. Biochimica et Biophysica Acta 1401, 93-103.

Seack, J., Pancer, Z., Müller, I.M, Müller, W.E.G., 1997. Molecular cloning and primary structure of a Rhesus (Rh)-like protein from the marine sponge Geodia cydonium. Immunogentics 46, 493-498.

Soest, R.W.M. van, 1994. Demosponge distribution patterns. Soest RWM van, Balkema AA (eds) Sponges in time and space. Brookfield, Rotterdam, pp 213-223.

Stierle, A.C., Cardellina, J.H.I., Singleton, F.L., 1988. A marine *Micrococcus* produces metabolites ascribed to the sponge *Tedania ignis*. Experientia 44, 1021.

Sueoka, N., 1959. A statistical analysis of deoxyribonucleic acid distribution in density gradient centrifugation. Proc. Natl. Acad. Sci. USA, 45, 1480-1490.

Sueoka, N., 1961. Correlation between base composition of the deoxyribonucleic acid and amino acid and composition of proteins. Proc. Natl. Acad. Sci. USA, 47, 1141-1149.

Sueoka, N., Marmur, J., Doty, P., 1959. Heterogeneity in deoxyribonucleic acids: II. Dependence of the density of deoxyribonucleic acids on guanine-cytosine content. Nature 183, 1429-1433.

Tachibana, K., Scheuer, P.J., Tsukitani, Y., Kikushi, H., Engen, D.V., Clardy, J., Gopichand, Y., Schmitz, F.J., 1981. Okadaic acid a cytotoxic polyether from the marine sponges of the genus *Halichondria*. J. Am. Chem. Soc. 103, 2469-2471.

Thiery, J.P., Macaya, G., Bernardi, G., 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J. Mol. Biol. 108, 219-235.

Ueda, T., Suga, Y., Matsuguchi, T., 1995. Molecular phylogenetic analysis of a soil microbial community. Eur. J. Soil. Sci. 46, 415-421.

Vacelet, J., 1970. Description de cellules à bactèries intranuclèaires chez des èponges *Verongia*. J. Microscopy 9, 333-346.

Vacelet, J., 1971. Ètude en microscopie èlectronique de l'association entre une cyanophycèe chroococcale et une èponge du genre *Verongia*. J Microscopy 12, 363-380.

Vacelet, J., 1975. Etude en microscopie electronique de l'association entre bacteries et Spongiaires du genre *Verongia*. J. Microscopie, T. 23, 271-288.

Vacelet, J., Boury-Esnault, N., 1995. Carnivorous sponges. Nature 373, 333-335.

Vacelet, J., Donadey, C., 1977. Electron microscope study of the association between some sponges and bacteria. J. Exp. Mar. Biol. Ecol. 30, 301-314.

Vacelet, J., Vacelet, E., Gaino, E., Gallissian, M.F., 1994. Bacterial attack of sponging skeleton during the 1986-1990 Mediterranean sponge disease. In: Van Soest R.W.M., Van Kempen, T.M.G, Braekman JC (eds) Sponges in time and space. Balkema, Rotterdam, 355-362.

Vatova, A., 1928. Compendio della flora e fauna del Mare Adriatico presso Rovigno. C. Ferrari, Venezia.

Vogel, S., 1977. Current-induced flow through living sponges in nature. Proc. Natl. Acad. Sci USA 74, 2069-2071

Wagner, C., Steffen, R., Koziol, C., Batel, R., Lacorn, M., Steinhart, H., Simat, T., Müller, W.E.G., 1998. Apoptosis in marine sponges: a biomarker for environmental stress (cadmium and bacteria). Marine Biology 131, 411-421.

Webb, V.L., Maas, E., 2002. Sequence analysis of 16S rRNA gene of cyanobacteria associated with the marine sponge Mycale (Carmia) hentscheli. FEMS Microbiology Letters 207, 43-47.

Webster, N.S., Webb, R.I., Ridd, M.J., Hill, R.T., Negri, A.P., 2001. The effect of copper on the microbial community of a coral ref sponge. Environ. Microbiol. 3(1), 19-31.

Weinbaum, G., Burger, M.M., 1973. A two-component system for surface guided reassociation of animal cells. Nature 244: 510-512.

Wiens, M., Koziol, C., Batel, R., Müller, W.E.G., 1998. Phenylalanine hydroxylase from the sponge *Geodia cydonium*: implication for allorecognition and evolution of aromatic amino acic hydroxylase. Developmental and comparative immunology 22 (5/6), 469-478.

Wiens, M., Koziol, C., Batel, R., Müller, W.E.G., 1999. Prolidase in the marine sponge Suberites domuncula: enzyme activity, molecular cloning, and phyloenetic relationship. Mar. Biotechnol. 1, 191-199.

Wiens, M., Krasko, A., Müller, I.M., Müller, W.E.G., 2000. Molecular evolution of apoptotic pathways: cloning of key domains from sponges (Bcl-2 homology domains and death domains) and their phylogenetic relationship. J. Mol. Evol. 50, 520-531.

Wiens, M., Krasko, A., Perovic, S., Müller, W.E.G., 2003. Caspase-mediated apoptosis in sponges: cloning and function of the phylogenetic oldest apoptotic proteases from Metazoa. Biochimica et Biophysica Acta, 1593, 179-189.

Wiens, M., Kuusksalu, A., Kelve, M., Müller, W.E.G.,1999. Origin of the interferon-inducibile (2'-5')oligoadenylate synthetases: cloning of the (2'-5')oligoadenylate synthetase from the marine sponge *Geodia cydonium*. FEBS letters 462, 12-18.

Wiens, M., Luckacs, B., Brümmer, F., Shokry, M., Ammar, A., Steffen, R., Batel, R., Diehl-Seifert, B., Schröeder, H.C., Müller, W.E.G., 2003. Okadaic acid: a potential defense toxin for the sponge *Suberites domuncula*. Mar. Biol. 142, 213-223.

Wilkinson, C.R., 1978. Microbial associations in sponges. II. Numerical analysis of sponge and water bacterial populations. Mar. Biol. 49, 169-176.

Wilkinson, C.R., Garrone, R., Vacelet, J., 1984. Marine sponges discriminate between food bacteria and bacterial symbionts: electron microscope radioautography and *in situ* evidence. Proc. R. Soc. Lond. B 220, 519-528.

Willmer, P., 1994. Invertebrate relationship. Cambridge University Press, Cambridge.

Wimmer, W., Perovic, S., Kruse, M., Schröeder, H.C., Krasko, A., Batel, R., Müller, W.E.G., 1999. Origin of the integrin-mediated signal transduction. Functional studies with cell cultures from the sponge *Suberites domuncula*. Eur. J. Biochem. 260, 156-165.

Yu, J. et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79-92.

Zerial M., Salinas, J., Filipski, J., Bernardi, G., 1986. Gene distribution and nucleotide sequence organization in the human genome. Eur. J. Biochem. 160, 479-85.

Zoubak, S., Clay, O., Bernardi, G., 1996. The gene distribution of the human genome. Gene 174, 95-102.

Appendix A

Alignment of the 24 sequences of bacterial clones showed in Tables 3.3 and 3.4.

```
Multalin version 5.4.1
Copyright I.N.R.A. France 1989, 1991, 1994, 1996
Published research using this software should cite
Multiple sequence alignment with hierarchical clustering
F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890
Symbol comparison table: blosum62
Gap weight: 12
Gap length weight: 2
Consensus levels: high=90% low=50%
Consensus symbols:
 ! is anyone of IV
 $ is anyone of LM
 % is anyone of FY
 # is anyone of NDQEBZ

 MSF:   1502    Check:   0         ..
 Name: SdB3           Len:  1502  Check: 8906  Weight:  2.59
 Name: SdB4           Len:  1502  Check: 9834  Weight:  0.68
 Name: GcB3           Len:  1502  Check: 3559  Weight:  0.68
 Name: SdB11          Len:  1502  Check: 6669  Weight:  0.82
 Name: SdB12          Len:  1502  Check: 3282  Weight:  0.94
 Name: SdB5           Len:  1502  Check: 2169  Weight:  0.71
 Name: GcB13          Len:  1502  Check: 8100  Weight:  0.71
 Name: GcB12          Len:  1502  Check: 5810  Weight:  0.94
 Name: GcB4           Len:  1502  Check: 6893  Weight:  1.18
 Name: SdB10          Len:  1502  Check: 9106  Weight:  0.68
 Name: GcB10          Len:  1502  Check: 9570  Weight:  0.68
 Name: GcB9           Len:  1502  Check: 5244  Weight:  0.85
 Name: GcB8           Len:  1502  Check: 1742  Weight:  1.15
 Name: GcB11          Len:  1502  Check: 8866  Weight:  1.15
 Name: SdB7           Len:  1502  Check: 3604  Weight:  0.71
 Name: SdB15          Len:  1502  Check: 2596  Weight:  0.71
 Name: SdB9           Len:  1502  Check:  992  Weight:  0.52
 Name: SdB13          Len:  1502  Check: 9399  Weight:  0.52
 Name: SdB14          Len:  1502  Check: 8320  Weight:  0.66
 Name: GcB6           Len:  1502  Check: 7265  Weight:  0.68
 Name: SdB8           Len:  1502  Check: 9068  Weight:  1.01
 Name: GcB7           Len:  1502  Check: 1366  Weight:  1.18
 Name: SdB6           Len:  1502  Check: 5364  Weight:  1.79
 Name: GcB5           Len:  1502  Check: 6440  Weight:  2.45
 Name: Consensus      Len:  1502  Check: 4529  Weight:  0.00

//

                    1                                                  50
              SdB3  .......... .......... .......... .......... ..........
              SdB4  .......... .......... .......... .......... ...GGCTTGA
              GcB3  .......... .......... .......... .......... ..........
             SdB11  .......... .......... .......... .......... ..........
             SdB12  .......... .......... .......... .......... ...GGCTTGA
              SdB5  .......... .......... .......... .......... ...GGCTTGA
             GcB13  .......... .......... .......... .......... ..........
             GcB12  .......... .......... .......... .......... ........GA
              GcB4  .......... .......... .......... .......... ..CGGCTTGA
             SdB10  .......... .......... .......... .......... ..........
             GcB10  .......... .......... .......... .......... ..........
```

```
          GcB9  .......... .......... .......... .......... ..........
          GcB8  .......... .......... .......... .......... ...GGCTTGA
         GcB11  .......... .......... .......... .......... ...GGCTTGA
          SdB7  .......... .......... .......... .......... ......TTGA
         SdB15  .......... .......... .......... .......... ...GGCTTGA
          SdB9  .....CGGAT CCACTAGTAA CG.CCGCCAG TGTGCTGGAA TTCGGC..GA
         SdB13  .......... .......... .......... .......... ...GGCTTGA
         SdB14  .......... .......... .......... .......... ...GGCTTGA
          GcB6  .......... .......... .G.CCACCAG TGTGCTGGAA TTCGGC..GA
          SdB8  .......... .......... .......... .......... ........GA
          GcB7  .......... .......... ...CCGCCAG TGTGCTGGAA TTCGGC..GA
          SdB6  .......... .......... .......... .......... ..........
          GcB5  GAGCTCGGAT CCACTAGTAA CGGCCGCCAG TGTGCTGGA. .TCGGCTTGA
     Consensus  .......... .......... .......... .......... ...ggc..ga

                51                                                  100
          SdB3  .......... .......... .......... .......... ..........
          SdB4  GTT.GATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
          GcB3  .......... ..CTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
         SdB11  ....GATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
         SdB12  GTTTGATCCT GGCTCAGATT GAACGCTGGC GGCACGCTTT ACACATGCAA
          SdB5  GTT.GATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
         GcB13  ....GATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
         GcB12  GTTTGATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
          GcB4  GTTTGATCCT GGCTCAGGAT GAACGCTGGC GGCGTGCCTA ATACATGCAA
         SdB10  GTTTGATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
         GcB10  ...TGATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
          GcB9  ..TTGATCCT GGCTCAGATT GAACGCTGGC GGCAGGCCTA ACACATGCAA
          GcB8  GTTTGATCCT GGCTCAGAAC GAACGCTGGC GGCAGGCCTA ACACATGCAA
         GcB11  GTTTGATCCT GGCTCAGAAC GAACGCTGGC GGCAGGCCTA ACACATGCAA
          SdB7  GTT.GATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
         SdB15  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
          SdB9  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
         SdB13  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
         SdB14  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
          GcB6  GTTTGATCCT GGCTCAGGAT GAACGCTGGC GGCGTGCCTA ATACATGCAA
          SdB8  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
          GcB7  GTTTGATCCT GGCTCAGGAC GAACGCTGGC GGCGTGCCTA ATACATGCAA
          SdB6  .......... .......... .......... .......... ..........
          GcB5  GTTTGATCCT GGCTCAGAAC GAACGCTGGC GGCAGGCCTA ACACATGCAA
     Consensus  gtttgatcct ggctcaga.. gaacgctggc ggca.gccta acacatgcaa

                101                                                 150
          SdB3  .......... .......... .......... .......... ..........
          SdB4  GTCGAGCGGT AACAGAAAGT AG....CTT. .....GC.TA CTTTGCTGAC
          GcB3  GTCGAGCGGT AACAGAGAGT AG....CTT. .....GC.TA CTTTGCTGAC
         SdB11  GTCGAGCGGA AACGAAGAGT AG....CTT. .....GC.TA CTCTGGCGTC
         SdB12  GTCGAGCGGC AGCGCAGGGG TG....CTT. .....GC.AC CCTTGGCGGC
          SdB5  GTCGAGCGGA AACGAGTTAA CTGACCCTTC GGGTGACGTT AACGG.CGTC
         GcB13  GTCGAGCGGA AACGACACTA ACAATCCTTC GGGT.ACGTT AATGGGCGTC
         GcB12  GTCGAGCGGT AACAGAAAGA AAG...CTT. .....GCTTT CTTTGCTGAC
          GcB4  GTCGAGCGAA CAGATAAGGA G.....CTT. .....GC.TC CTTTGACGTT
         SdB10  GTCGAGCGGT AACAGGACTA G.....CTT. .....GC..T AGTTGCTGAC
         GcB10  GTCGAGCGGT AACAGGACTA G.....CTT. .....GC..T AGTTGCTGAC
          GcB9  GTCGAGCGGT AACAGGACTA G.....CTT. .....GC..T AGTTGCTGAC
          GcB8  GTCGAACGGA .......... .........T CCTTCGGGAT
         GcB11  GTCGAACGGA .......... .........T CCTTCGGGAT
          SdB7  GTCGAGCGGA CAGAA.GGGA G.....CTT. .....GC..T CCC.GGATGT
         SdB15  GTCGAGCGAA TCAAT.GGGA G.....CTT. .....GC..T CCC.TGAGAT
```

```
      SdB9  GTCGAGCGGA G.ATTTGGGA G.....CTT. .....GC..T CCCAA.ATCT
     SdB13  GTCGAGCGGA G.ATTTGGGA G.....CTT. .....GC..T CCCAA.ATCT
     SdB14  GTCGAGCGGA TCAATGGGGA G.....CTT. .....GC..T CCCCTGAGAT
      GcB6  GTCGAGCGAA T.GATGAGGA G.....CTT. .....GC..T CCTCTGAT.T
      SdB8  GTCGAGCGAA TCTGA.GGGA G.....CTT. .....GC..T CCCAA.AGAT
      GcB7  GTCGAGCGGA G.AATTGGGA G.....CTT. .....GC..T CCCAA.TTCT
      SdB6  .......... .......... .......... .......... ..........
      GcB5  GTCGAACGGA .......... .......... .........T CCTTCGGGAT
 Consensus  gtcgagcgga .......... ......ctt. .....gc..t c.tt...g..

            151                                                 200
      SdB3  .......... .......... .......... .......... ..........
      SdB4  GAGCGGCGGA CGGGTGAGTA ATGCTTGGG. AACATGCCTT GAGGTGGGGG
      GcB3  GAGCGGCGGA CGGGTGAGTA ATGCTTGGG. AACATGCCTT GAGGTGGGGG
     SdB11  GAGCGGCGGA CGGGTGAGTA ATGCTTGGG. AAGCTACCTA GTCGAGGGGG
     SdB12  GAGCGGCGGA CGGGTGAGTA ATGCATGGG. AATATGCCTA GTAGTGGGGA
      SdB5  GAGCGGCGGA CGGGTGAGTA ATGCCTGGG. AATATGCCTT GATGTGGGGG
     GcB13  GAGCGGCGGA CGGGTGAGTA ATGCCTAGG. AAATTGCCTT GATGTGGGGG
     GcB12  GAGCGGCGGA CGGGTGAGTA ATGCCTAGG. GATCTGCCCA GTCGAGGGGG
      GcB4  .AGCGGCGGA CGGGTGAGTA ACACGTGGAT AACCTACCTA TAAGACTGGG
     SdB10  AAGCGGCGGA CGGGTGCGTA ACACGTGGG. AATCTGCCCG GTAGTGGGGG
     GcB10  AAGCGGCGGA CGGGTGCGTA ACACGTAGG. AATCTGCCCG GTAGTGGGGG
      GcB9  AAGCGGCGGA CGGGTGCGTA ACACGTAGG. AATCTGCCCG GTAGTGGGGG
      GcB8  TAGTGGCAGA CGGGTGAGTA ACGCGTGGG. AAGCTACCTT GTGGTAGGGA
     GcB11  TAGTGGCAGA CGGGTGAGTA ACGCGTGGG. AAGCTACCTT GTGGTAGGGG
      SdB7  TAGCGGCGGA CGGGTGAGTA ACACGTGGGT AACCTGCCTG TAAGACTGGG
     SdB15  TAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTGCCTA TAAGACTGGG
      SdB9  TAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTGCCCT GCAGACTGGG
     SdB13  TAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTGCCCT GCAGACTGGG
     SdB14  CAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTACCTA TAAGACTGGG
      GcB6  TAGCGGCGGA CGGGTGAGTA ACACGTGGGT AATCTGCCTG TAAGACGGGG
      SdB8  TAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTGCCTG TAAGACTGGG
      GcB7  TAGCGGCGGA CGGGTGAGTA ACACGTGGGC AACCTGCCCT GCAGACTGGG
      SdB6  .......... .......... .......... .......... ..........
      GcB5  TAGTGGCAGA CGGGTGAGTA ACGCGTGGG. AAGCTACCTT GTGGTAGGGG
 Consensus  .agcggcgga cgggtgagta ac.cgtggg. aa.ctgcct. g..g..gggg

            201                                                 250
      SdB3  .......... .......... .......... .......... ..........
      SdB4  ACAACAGTTG GAAACGACTG CTAATACCGC ATAA...... .TGTCT.ACG
      GcB3  ACAACAGTTG GAAACGACTG CTAATACCGC ATAA...... .TGTCT.ACG
     SdB11  ACAACCATTG GAAACGATGG CTAATACCGC ATAC...... .GCCCT.ACG
     SdB12  ATAACTTTGG GAAACCAGAG CTAATACCGC ATAC...... .GCTCT.ACG
      SdB5  ATAACCATTG GAAACGATGG CTAATACCGC ATAA...... .CGCCT.TCG
     GcB13  ATAACCATTG GAAACGATGG CTAATACCGC ATGA...... .TGCCT.ACG
     GcB12  ATAACAGTTG GAAACGACTG CTAATACCGC ATAC...... .GCCCT.ACG
      GcB4  ATAACTTCGG GAAACCGGAG CTAATACCGG ATAACATATT GAACCTCATG
     SdB10  ATAGCCCGGA GAAATTCGGA TTAATACCGC ATAC...... .GCCCT.AAG
     GcB10  ATAGCCCGGA GAAATTCGGA TTAATACCGC ATAC...... .GCCCT.AAG
      GcB9  ATAGCCCGGA GAAATTCGGA TTAATACCGC ATAC...... .GCCCT.AAG
      GcB8  ACAACAGTTG GAAACGACTG CTAATACCCT ATGA...... .GCCCT.AAG
     GcB11  ACAACAGTTG GAAACGACTG CTAATACCCT ATGA...... .GCCCT.ATG
      SdB7  ATAACTCCGG GAAACCGGAG CTAATACCGG ATAGTTCCTT GAACCGCATG
     SdB15  ATAACTTCGG GAAACCGGAG CTAATACCGG ATACGTTCTT TTCTCGCATG
      SdB9  ATAACTCCGG GAAACCGGAG CTAATACCGG GTAATACATC GCACCGCATG
     SdB13  ATAACTCCGG GAAACCGGAG CTAATACCGG GTAATACATC GCACCGCATG
     SdB14  ATAACTCCGG GAAACCGGGG CTAATACCGG ATAACATTTT CCACTGCATA
      GcB6  ATAACTCCGG GAAACCGGGG CTAATACCGG ATAACAAGAG AAGAAGCATT
      SdB8  ATAACTCCGG GAAACCGGGG CTAATACCGG ATAATATCTA TTTATACATA
```

```
      GcB7  ATAACTCCGG GAAACCGGAG CTAATACCGG GTAATACATC GCACCGCATG
      SdB6  .......... .AAACGGCTG CTAATACCGC ATAC...... .GCCCT.ACG
      GcB5  ACAACAGTTG GAAACGACTG CTAATACCCT ATGA...... .GCCCT.ATG
 Consensus  ataac....g gaaac....g ctaataccg. ataa...... ..ccct.a.g

            251                                                 300
      SdB3  .......... .......... .......... .......... ..........
      SdB4  GACCAAAGGG GG......CT TCG..G..CT CTCGCCTTTA GATTGGCCCA
      GcB3  GACCAAAGGG GG......CT TCG..G..CT CTCGCCTTTA GATTGGCCCA
     SdB11  GGGGAAAGGA GGGGAC..CT TCG..GGCCT TTCGCGATTA GATGTGCCCA
     SdB12  GAGGGAAGCG GGGGAT..CT TTT..GACCT CGCGCTATTA GAGTAGCCCA
      SdB5  GGCCAAAGAG GGGGAT..CT TCG..GACCT CTCGCGTCAA GATTAGCCCA
     GcB13  GGCCAAAGAG GGGGAC..CT TCG..GGCCT CTCGCGTCAA GATATGCCTA
     GcB12  GGGGAAAGGA GGGGAC..CT TCG..GGCCT TTCGCGATTG GATGAACCTA
      GcB4  GTTCAATAGT GAAAGG..CG GCT..TTGCT GTCACTTATA GATGGATCCG
     SdB10  GGGGAAAGAT GGCCTCTTCT TGA..AAGCT ATCACTATCC GATGAGCCTG
     GcB10  GGGGAAAGAT GGCCTCTTCT TGA..AAGCT ATCACTATCG GATGAGCCTG
      GcB9  GGGGAAAGAT GGCCTCTTCT TGA..AAGCT ATCACTATCG GATGAGCCTG
      GcB8  GGGGAAAGAT .......... ........TT ATCGCCATGA GATGTGCCCG
     GcB11  GGGGAAAGAT .......... ........TT ATCGCCATGA GATGTGCCCG
      SdB7  GTTCAAGGAT GAAAGACGGT TTC...GGCT GTCACTTACA GATGGACCCG
     SdB15  AGAGAAGATG GAAAGACGGT TTA...CGCT GTCACTTATA GATGGGCCCG
      SdB9  GTGCAATGTT GAAAGTTGGC TTTC.GAGCT AACACTGCAG GATGGGCCCG
     SdB13  GTGCAATGTT GAAAGTTGGC TTTCTGAGCT AACACTGCAG GATGGGCCCG
     SdB14  GTGGAGAATT AAAAGATGGC TTC...GGCT ATCACTTACA GATGGGCCCG
      GcB6  TCTTCTTTTT GAAAGTCGGC ATCT..CGCT GACACTTACA GATGAGCCCG
      SdB8  TAATTAGATT GAAAGATGGT TCT....GCT ATCACTTACA GATGGGCCCG
      GcB7  GTGCAATGTT GAAAGTTGGC TTTC.GAGCT AACACTGCAG GATGGGCCCG
      SdB6  GGGGAAAGCA GGGGAT..CT TCG..GACCT TGCGCTATTG GATGAGCCTA
      GcB5  GGGGAAAGAT .......... ........TT ATCGCCATGA GATGTGCCCG
 Consensus  ggggaaag.t g......... t.......ct .tcgct.t.a gatg.gcccg

            301                                                 350
      SdB3  .......... .......... .......... .......... ..........
      SdB4  AGTGGGATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CAACGATCCC
      GcB3  AGTGGGATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CAACGATCCC
     SdB11  AGTGGGATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CGACGATCCC
     SdB12  TGTCCGATTA GCTAGTTGGA GGGGTAACAG CCCACCAAGG CGATGATCGG
      SdB5  GGTGGGATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CGACGATCCC
     GcB13  GGTGGGATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CGACGATCCC
     GcB12  GGTGGGATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATCCC
      GcB4  CGCCGTATTA GCTAGTTGGT AAGGTAACGG CTTACCAAGG CAACGATACG
     SdB10  CGTCGGATTA GCTAGTTGGT GGGGTAAAGG CCTACCAAGG CAACGATCCG
     GcB10  CGTCGGATTA GCTAGTTGGT GGGGTAAAGG CCTACCAAGG CAACGATCCG
      GcB9  CGTCGGATTA GCTAGTTGGT GGGGTAAAGG CCTACCAAGG CAACGATCCG
      GcB8  CGTTAGATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATCTA
     GcB11  CGTTAGATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATCTA
      SdB7  CGGCGCATTA GCTAGTTGGT GAGGTAACGG CTCACCAAGG CGACGATGCG
     SdB15  CGGCGCATTA GCTAGTTGGT GAGGTAATGG CTCACCAAGG CGACGATGCG
      SdB9  CGGCGCATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATGCG
     SdB13  CGGCGCATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATGCG
     SdB14  CGGCGCATTA GCTAGTTGGT GAGGTAAGGG CTCACCAAGG CGACGATGCG
      GcB6  CGGCGCATTA GCTAGTTGGT GAGGTAACGG CTCACCAAGG CGACGATGCG
      SdB8  CGGCGCATTA GCTAGTTGGT GAGGTAACGG CTCACCAAGG CGACGATGCG
      GcB7  CGGCGCATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATGCG
      SdB6  AGTCGGATTA GCTAGTTGGT GAGGTAAAGG CTCACCAAGG CGACGATCCG
      GcB5  CGTTAGATTA GCTAGTTGGT AAGGTAATGG CTTACCAAGG CGACGATCTA
 Consensus  cgtcggatta gctagttggt gaggtaatgg ct.accaagg cgacgatccg
```

```
           351                                                        400
     SdB3  .......... .......... .......... .......... ..........
     SdB4  TAGCTGGTTT GAGAGGATGA TCAGCCACAC TGGAACTGAG ACACGGTCCA
     GcB3  TAGCTGGTTT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
    SdB11  TAGCTGGTTT GAGAGGATGA TCAGCCACAC TGGAACTGAG ACACGGTCCA
    SdB12  TAGCTGGTCT AAGAGGATGA TCAGCCACAC CGGGACTGAG ACACGGCCCG
     SdB5  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGAACTGAG ACACGGTCCA
    GcB13  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGAACTGAG ACACGGTCCA
    GcB12  TAGCTGTTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
     GcB4  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGAACTGAG ACACGGTCCA
    SdB10  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
    GcB10  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
     GcB9  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
     GcB8  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
    GcB11  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
     SdB7  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
    SdB15  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
     SdB9  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
    SdB13  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
    SdB14  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
     GcB6  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
     SdB8  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
     GcB7  TAGCCGACCT GAGAGGGTGA TCGGCCACAC TGGGACTGAG ACACGGCCCA
     SdB6  TAGCTGGTTT GAGAGGATGA TCAGCCACAT CGGGACTGAG ACACGGCCCG
     GcB5  TAGCTGGTCT GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA
Consensus  tagctggtct gagaggatga tcagccacac tgggactgag acacggccca

           401                                                        450
     SdB3  .......... .......... .......... .......... ....CAAGCC
     SdB4  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGAAAGCC
     GcB3  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
    SdB11  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
    SdB12  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGGACAATG GGGGCAACCC
     SdB5  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
    GcB13  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGAAAGCC
    GcB12  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGGGAAACCC
     GcB4  GACTCCTACG GGAGGCAGCA GTAGGGAATA TTGCACAATG GGCGCAAGCC
    SdB10  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
    GcB10  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
     GcB9  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGCACAATG GGCGCAAGCC
     GcB8  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGGACAATG GGGGCAACCC
    GcB11  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGGACAATG GGGGCAACCC
     SdB7  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
    SdB15  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
     SdB9  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
    SdB13  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
    SdB14  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCGGCAATG GGCGAAAGCC
     GcB6  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCGGCAATG GGCGAAAGCC
     SdB8  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
     GcB7  GACTCCTACG GGAGGCAGCA GTAGGGAATC TTCCGCAATG GACGAAAGTC
     SdB6  AACTCCTACG GGAGGCAGCA GTGGGGAATA TTGGACAATG GGGGCAACCC
     GcB5  GACTCCTACG GGAGGCAGCA GTGGGGAATA TTGGACAATG GGGGCAACCC
Consensus  gactcctacg ggaggcagca gtggggaata ttgcacaatg ggcgcAAgcC

           451                                                        500
     SdB3  TGATGCAGCC ATGCCGCGTG TATGAAGAAG GCCTTCGGGT TGTAAAGTAC
     SdB4  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCCTTCGGGT TGTAAAGCAC
     GcB3  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCCTTCGGGT TGTAAAGCAC
    SdB11  TGATGCAGCC ATGCCGCGTG TATGAAGAAG GCCTTCGGGT TGTAAAGTAC
```

```
      SdB12  TGATCCAGCG ATGCCGCGTG AGTGAAGAAG GCTCTCGGGT TGTAAAGCTC
       SdB5  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCCTTCGGGT TGTAAAGCAC
      GcB13  TGATGCAGCC ATGCCGCGTG TATGAAGAAG GCCTTCGGGT TGTAAAGTAC
      GcB12  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCCTTAGGGT TGTAAAGCAC
       GcB4  TGATGCAG.. .....GCGTG TGTGAAGAAG GCTCTAGGGT TGTAAAGCAC
      SdB10  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCTCTAGGGT TGTAAAGCAC
      GcB10  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCTCTAGGGT TGTAAAGCAC
       GcB9  TGATGCAGCC ATGCCGCGTG TGTGAAGAAG GCTCTAGGGT TGTAAAGCAC
       GcB8  TGATCCAGCC ATGCCGCGTG TGTGATGACG GCCTTAGGGT TGTAAAGCAC
      GcB11  TAATCCAGCC ATGCCGCGTG TGTGATGACG GCCTTAGGGT TGTAAAGCAC
       SdB7  TGACGGAGCA ACGCCGCGTG AGTGATGAAG GTTTTCGGAT CGTAAAGCTC
      SdB15  TGACGGAGCA ACGCCGCGTG AACGAAGAAG GCCTTCGGGT CGTAAAGTTC
       SdB9  TGACGGAGCA ACGCCGCGTG AGTGACGAAG GCCTTCGGGT CGTAAAGCTC
      SdB13  TGACGGAGCA ACGCCGCGTG AGTGACGAAG GCCTTCGGGT CGTAAAGCTC
      SdB14  TGACCGAGCA ACGCCGCGTG AGCGATGAAG GCCTTCGGGT CGTAAAGCTC
       GcB6  TGACCGAGCA ACGCCGCGTG AGCGATGAAG GCCTTCGGGT CGTAAAGCTC
       SdB8  TGACGGAGCA ACGCCGCGTG AGTGATGAAG GTTTTCGGAT CGTAAAACTC
       GcB7  TGACGGAGCA ACGCCGCGTG AGTGACGAAG GCCTTCGGGT CGTAAAGCTC
       SdB6  TGATCCAGCC ATGCCGCGTG TGTGAAGAAG GCTTTCGGGT TGTAAAGCAC
       GcB5  TGATCCAGCC ATGCCGCGTG TGTGATGACG GCCTTAGGGT TGTAAAGCAC
  Consensus  TGAtgcAGCc AtGCCGCGTG tgTGAaGAaG GCctTcGGGT tGTAAAGcaC

             501                                                 550
       SdB3  TTTCAGTTGT GAGGAA.GGG GGTGTCGTTA ATAGCGGCAT CTCTTGACGT
       SdB4  TTTCAGTCAG GAGGAA.AGG GTGTGAGTTA ATACCTCACA TCTGTGACGT
       GcB3  TTTCAGTCAG GAGGAA.AGG TTAGTAGTTA ATACCTGCTA GCTGTGACGT
      SdB11  TTTCAGCGAG GAGGAA.AGG TTAGTAGCTA ATAACTGCTA GCTGTGACGT
      SdB12  TTTCGCAGGG AAAGAA.ACG GCAATGGTAA ATAGCTATTG CAACTGACGG
       SdB5  TTTCAGTCGT GAGGAA.GGT GGTGTAGTTA ATAGCTGCAT TATTTGACGT
      GcB13  TTTCAGTTGT GAGGAA.GGG TGTGTAGTTA ATAGCTGCGC ATCTTGACGT
      GcB12  TTTCAGTAGG GAGGAA.AGG TAATGGCTTA ATACGCTATT ACTGTGACGT
       GcB4  TTTCAGCGAG GAGGAA.AGG TTAGTAGTTA ATACCTGCTA GCTGTGACGT
      SdB10  TTTCAGCGAG GAGGAA.AGG TTGAAGATTA ATACTCTTTA GCTGTGACGT
      GcB10  TTTCAGCGAG GAGGAA.AGG TTGAAGATTA ATACTCTTTA GCTGTGACGT
       GcB9  TTTCAGCGAG GAGGAA.AGG TTGAAGATTA ATACTCTTTA GCTGTGACGT
       GcB8  TTTCAGCAGT GAAGAT.AA. .......... .......... ....TGACAT
      GcB11  TTTCAGCAGT GAAGAT.AA. .......... .......... ....TGACAT
       SdB7  TGTTGTTAGG GAAGAACAAG TGCAAGAGTA ACTGCT.TGC ACCTTGACGG
      SdB15  TGTTGTTAGG GAAGAACAAG TACCAGAGTA ACTGCT.GGT ACCTTGACGG
       SdB9  TGTTGTTAGG GAAGAACAAG TACCGTTCGA ATAGGGCGGT ACCTTGACGG
      SdB13  TGTTGTTAGG GAAGAACAAG TACCGTTCGA ATAGGGCGGT ACCTTGACGG
      SdB14  TGTTGTTAGG GAAGAACAAG TACCGTTCAA ACAGGGCGGT ACCTTGACGG
       GcB6  TGTTGTTAGA GAAGAACAAG TACGAGAGTA ACTGCTCG.T ACCTTGACGG
       SdB8  TGTTGTTAGG GAAGAACAAG TATCGGAGTA ACTGCC..GT ACCTTGACGG
       GcB7  TGTTGTTAGG GAAGAACAAG TACCGTTCGA ATAGGGCGGC ACCTTGACGG
       SdB6  TTTCAGCGAG GAGGAA.GGC TCTAAAGTTA ATACCTTTAG GGATTGACGT
       GcB5  TTTCAGCAGT GAAGATAA.. .......... .......... ....TGACAT
  Consensus  TtTcagt.gg GAgGAa.agg t......tta ata.c..... .c..TGACgt

             551                                                 600
       SdB3  TAGCAACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
       SdB4  TACTGACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
       GcB3  TACTGACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
      SdB11  TACTCGCAGA AGAA.GGACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
      SdB12  TACCCTGATA AGAA.GCACC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       SdB5  TAGCGACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
      GcB13  TAGCAACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
      GcB12  TACCTACAGA AGAA.GGACC GGCTAACTTC GTGCCAGCAG CCGCGGTAAT
       GcB4  TACTGACAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
```

```
      SdB10  TACTCGCAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
      GcB10  TACTCGCAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
       GcB9  TACTCGCAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
       GcB8  TAACTGCAGA AGAA.GCCCC GGCTAACTTC GTGCCAGCAG CCGCGGTAAT
      GcB11  TAACTGCAGA AGAA.GCCCC GGCTAACTTC GTGCCAGCAG CCGCGGTAAT
       SdB7  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
      SdB15  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       SdB9  TACCCTAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
      SdB13  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
      SdB14  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       GcB6  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       SdB8  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       GcB7  TACC.TAACC AGAAAGCCAC GGCTAACTAC GTGCCAGCAG CCGCGGTAAT
       SdB6  TACTCGCAGA AGAA.GCACC GGCTAACTCC GTGCCAGCAG CCGCGGTAAT
       GcB5  TAACTGCAGA AGAA.GCCCC GGCTAACTTC GTGCCAGCAG CCGCGGTAAT
  Consensus  TAcc..cAga AGAA.GCacC GGCTAACT.C GTGCCAGCAG CCGCGGTAAT

             601                                                 650
       SdB3  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGCATGCAG
       SdB4  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
       GcB3  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
      SdB11  ACGGAGGGTC CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
      SdB12  ACGGAGGGTC CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
       SdB5  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGCATGCAG
      GcB13  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGCATGCAG
      GcB12  ACGAGGGGTC CAAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
       GcB4  ACGGAGGGTG CGAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTACGCAG
      SdB10  ACGGAGGGTG CAAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTGCGTAG
      GcB10  ACGGAGGGTG CAAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTGCGTAG
       GcB9  ACGGAGGGTG CAAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGTGCGTAG
       GcB8  ACGAAGGGGG CTAGCGTTGT TCGGAATCAC TGGGCGTAAA GAGTACGTAG
      GcB11  ACGAAGGGGG CTAGCGTTGT TCGGAATCAC TGGGCGTAAA GAGTACGTAG
       SdB7  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GGGCTCGCAG
      SdB15  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
       SdB9  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
      SdB13  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
      SdB14  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
       GcB6  ACGTAGGTGG CAAGCGTTAT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
       SdB8  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
       GcB7  ACGTAGGTGG CAAGCGTTGT CCGGAATTAT TGGGCGTAAA GCGCGCGCAG
       SdB6  ACGGAGGGTG CAAGCGTTAA TCGGAATTAC TGGGCGTAAA GCGCGCGTAG
       GcB5  ACGAAGGGGG CTAGCGTTGT TCGGAATCAC TGGGCGTAAG ACTACGTAGG
  Consensus  ACGgAGGgtg C.AGCGTTaa tCGGAATtAc TGGGCGTAAa gcg..cgcaG

             651                                                 700
       SdB3  GCGGTCTGTT AAGCAAGATG TGAAAGCCCG GGGCTCAACC TCGGAACAGC
       SdB4  GCGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
       GcB3  GCGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
      SdB11  GCGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
      SdB12  GCGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
       SdB5  GTGGTTCGTT AAGTCAGATG TGAAAGCCCG GGGCTCAACC TCGGAACTGC
      GcB13  GTGGTTCATT AAGTCAGATG TGAAAGCCCG GGGCTCAACC TCGGAACTGC
      GcB12  GCGGTTCATT AAGCCAGATG TGAAATCCCC GGGCTCAACC TGGGAATTGC
       GcB4  GCGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
      SdB10  GCGGCTTGTT AAGTTGGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
      GcB10  GCGGCTTGTT AAGTTGGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
       GcB9  GCGGCTTGTT AAGTTGGATG TGAAAGCCCC GGGCTCAACC TGGGAACTGC
       GcB8  GCGGACTGAT AAGTTAGGGG TGAAATCCCA AGGCTCAACC TTGGAACTGC
      GcB11  GCGGACTGAT AAGTTAGGGG TGAAATCCCA AGGCTCAACC TTGGAACTGC
```

```
      SdB7  GCGGTTTCTT AAGTCTGATG TGAAAGCCCC CGGCTCAACC GGGGAGGGTC
     SdB15  GTGGTTCCTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
      SdB9  GCGGTCTTTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
     SdB13  GCGGTCTTTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
     SdB14  GCGGTCTCTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
      GcB6  GCGGTCTCTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
      SdB8  GCGGTTCTTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
      GcB7  GCGGTCTTTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC GTGGAGGGTC
      SdB6  GTGGTTTGTT AAGCGAGATG TGAAAGCCCC GGGCTTAACC TGGGAACGGC
      GcB5  GCGGACTGAT AAGTTAGGGG TGAAATCCCA GGGCTCAACC TTGGAACTGC
 Consensus  GcGGtttgtT AAGt.aGatG TGAAAgCCC. gGGCTCAACC t.GGAactgC

            701                                                750
      SdB3  ATTTTGAACT GGCAGACTAG AGTCTTGTAG AGGGGGGTAG AATTTCAGGT
      SdB4  ATTTCGAACT GGCAAACTAG AGTGTGATAG AGGGTGGTAG AATTTCAGGT
      GcB3  ATTTCGAACT GGCAAACTAG AGTGTGATAG AGGGTGGTAG AATTTCAGGT
     SdB11  ATTTCGAACT GACAAACTAG AGTTTTGTAG AGGGTGGTAG AATTTCAGGT
     SdB12  ATTTCGAACT GACAAACTAG AGTTTTGTAG AGGGTGGTAG AATTTCAGGT
      SdB5  ATTTGAAACT GGCGGACTAG AGTACTGTAG AGGGGGGTAG AATTTCAGGT
     GcB13  ATTTGAAACT GGTGAACTAG AGTACTGTAG AGGGGGGTAG AATTTCAGGT
     GcB12  ATTTGGAACT GGTGAACTAG AGTCTTGTAG AGGGGGGTAG AATTTCAGGT
      GcB4  ATTTCGAACT GGCAAACTAG AGTGTGATAG AGGGTGGTAG AATTTCAGGT
     SdB10  ACCCAAAACT GACAAGCTAG AGTGCGGAAG AGGAGTGTGG AATTTCCTGT
     GcB10  ACCCAAAACT GACAAGCTAG AGTGCGGAAG AGGAGTGTGG AATTTCCTGT
      GcB9  ACCCAAAACT GACAAGCTAG AGTGCGGAAG AGGAGTGTGG AATTTCCTGT
      GcB8  CTTTGATACT GTCAGTCTTG AGATCGAGAG AGGTGAGTGG AACTCCGAGT
     GcB11  CTTTGATACT GTCAGTCTTG AGATCGAGAG AGGTGAGTGG AACTCCGAGT
      SdB7  ATTGGAAACT GGGAAACTTG AGTGCAGAAG AGGAGAGTGG AATTCCACGT
     SdB15  ATTGGAAACT GGGGAACTTG AGTGCAGAAG AGGAAAGTGG AATTCCAAGT
      SdB9  ATTGGAAACT GGAGGACTTG AGTGCAGAAG AGGAGAGTGG AATTCCACGT
     SdB13  ATTGGAAACT GGAGGACTTG AGTGCAGAAG AGGAGAGTGG AATTCCACGT
     SdB14  ATTGGAAACT GGGGGACTTG AGTACTGGAG AGGAGAGTGG AATTCCATGT
      GcB6  ATTGGAAACT GGGAGACTTG AGTGCAGGAG AGAAAAGTGG AATTCCACGT
      SdB8  ATTGGAAACT GGGGAACTTG AGTGCAGAAG AGGAAAGTGG AATTCCAAGT
      GcB7  ATTGGAAACT GGAGGACTTG AGTGCAGAAG AGGAGAGTGG AATTCCACGT
      SdB6  ATTTCGAACT GGCAAGCTAG AGTGTGGTAG AGGGTAGTGG AATTTCCTGT
      GcB5  CTTTGATACT GTCAGTCTTG AGATCGAGAG AGGTGAGTGG AACTCCGAGT
 Consensus  aTTtgaaACT GgcaaaCTaG AGt.c.g.AG AGG.gaGTgG AAtTtCa.GT

            751                                                800
      SdB3  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGGTG GCGAAGGCGG
      SdB4  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGATG GCGAAGGCAG
      GcB3  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGATG GCGAAGGCAG
     SdB11  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCAGTG GCGAAGGCGG
     SdB12  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCAGTG GCGAAGGCGG
      SdB5  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGGTG GCGAAGGCGG
     GcB13  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCAGTG GCGAAGGCGG
     GcB12  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGGTG GCGAAGGCGG
      GcB4  GTAGCGGTGA AATGCGTAGA GATCTGAAGG AATACCGATG GCGAAGGCAG
     SdB10  GTAGCGGTGA AATGCGTAG. TATAGGAAGG CACACCAGTG GCGAAGGCGA
     GcB10  GTAGCGGTGA AATGCGTAGA TATAGGAAGG CACACCAGTG GCGAAGGCGA
      GcB9  GTAGCGGTGA AATGCGTAGA TATAGGAAGG AACACCAGTG GCGAAGGCGA
      GcB8  GTAGAGGTGA AATTCGTAGA TATTCGGAAG CACACCAGTG GCGAAGGCGG
     GcB11  GTAGAGGTGA AATTCGTAGA TATTCGGAAG AACACCAGTG GCGAAGGCGG
      SdB7  GTAGCGGTGA AATGCGTAGA GATGTGGAGG AACACCAGTG GCGAAGGCGA
     SdB15  GTAGCGGTGA AATGCGTAGA GATTTGGAGG AACACCAGTG GCGAAGGCGA
      SdB9  GTAGCGGTGA AATGCGTAGA TATGTGGAGG AACACCAGTG GCGAAG.CGG
     SdB13  GTAGCGGTGA AATGCGTAGA TATGTGGAGG AACACCAGTG GCGAAGGCGG
     SdB14  GTAGCGGTGA AATGCGTAGA TATATGGAGG AACACCAGTG GCGAAGGCGG
```

```
      GcB6  GTAGCGGTGA AATGCGTAGA GATGTGGAGG AACACCAGTG GCGAAGGCGG
      SdB8  GTAGCGGTGA AATGCGTAGA GATTTGGAGG AACACCAGTG GCGAAGGCGA
      GcB7  GTAGCGGTGA AATGCGTAGA TATGTGGAGG AACACCAGTG GCGAAGGCGG
      SdB6  GTAGCGGTGA AATGCGTAGA TATAGGAAGG AACACCAGTG GCGAAGGCGG
      GcB5  GTAGAGGTGA AATTCGTAGA TATTCGGAAG AACACCAGTG GCGAAGGCGG
 Consensus  GTAGcGGTGA AATgCGTAGA gAT.tGaAgG aAcACCagTG GCGAAGGCgg

            801                                                 850
      SdB3  CCCCCTGGAC AAAGACTGAC GCTCAGATGC GAAAGCGTGG GGAGCAAACA
      SdB4  CCACCTGGGT CAACACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACG
      GcB3  CCACCTGGGT CAACACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACA
     SdB11  CCACCTGGAC AATAACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACA
     SdB12  CCACCTGGAC AATAACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACG
      SdB5  CCCCCTGGAC AGATACTGAC ACTCAGATGC GAAAGCGTGG GGAGCAAACA
     GcB13  CCCCCTGGAC AGATACTGAC ACTCAGATGC GAAAGCGTGG GGAGCAAACA
     GcB12  CCCCCTGGAC AAAGACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACA
      GcB4  CCACCTGGGT CAACACTGAC GCTCATGTAC GAAAGCGTGG GGAGCAAACA
     SdB10  CACTCTGGTC TGACACTGAC GCTGAGGTAC GAAAGCGTGG G.AGCAAACA
     GcB10  CACTCTGGTC TGACACTGAC GCTGAGGTAC GAAAGCGTGG GGAGCAAACA
      GcB9  CACTCTGGTC TGACACTGAC GCTGAGGTAC GAAAGCGTGG GGAGCAAACA
      GcB8  CTCACTGGCT CGATACTGAC GCTGAGGTAC GAAAGCGTGG GGAGCAAACA
     GcB11  CTCACTGGCT CGATACTGAC GCTGAGGTAC GAAAGCGTGG GGAGCAAACA
      SdB7  CTCTCTGGTC TGTAACTGAC GCTGAGGAGC GAAAGCGTGG GGAGCGAACA
     SdB15  CTTTCTGGTC TGTAACTGAC ACTGAGGCGC GAAAGCGTGG GGAGCAAACA
      SdB9  CTCTCTGGTC TGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
     SdB13  CTCTCTGGTC TGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
     SdB14  CTCTCTAGCC AGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
      GcB6  CTTTTTGGCC TGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
      SdB8  CTTTCTGGTC TGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
      GcB7  CTCTCTGGTC TGTAACTGAC GCTGAGGCGC GAAAGCGTGG GGAGCAAACA
      SdB6  CTACCTGGAC CAGCACTGAC ACTGAGGTGC GAAAGCGTGG GGAGCAAACA
      GcB5  CTCACTGGCT CGATACTGAC GCTGAGGTAC GAAAGCGTGG GGAGCAAACA
 Consensus  Ctc.CTGG.c .ga.ACTGAC gCTgAggtaC GAAAGCGTGG GGAGCAAACA

            851                                                 900
      SdB3  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTTGGAGGT
      SdB4  GGATTAGATA CCCCGGTAGT CCACGCCGTA AACGATGTCT ACTAGAAGCT
      GcB3  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTAGAAGCT
     SdB11  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTAGAAGCT
     SdB12  GGATTAGATA CCCCGGTAGT CCACGCCGTA AACGATGTCA ACTAGCCGAC
      SdB5  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTTGGAGGT
     GcB13  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTTGGAGGT
     GcB12  GGATTAGATA CCCTGGTGGT CCACGCCGTA AACGATGTCT ACTCGGAGTT
      GcB4  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTAGAAGCT
     SdB10  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAAT GCTAGTTGTC
     GcB10  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAAT GCTAGTTGTC
      GcB9  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCT ACTAGTCGTC
      GcB8  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAAT GCTAGTTGTC
     GcB11  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAAT GCTAGTTGTC
      SdB7  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAAGTGTT
     SdB15  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAAGTGTT
      SdB9  GGATTAGATA CCCTGGTAGT CCACGC.GTA AACGATGAGT GCTAGGTGTT
     SdB13  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAGGTGTT
     SdB14  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAGGTGTT
      GcB6  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAGGTGTT
      SdB8  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAAGTGTT
      GcB7  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAGT GCTAGGTGTT
      SdB6  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGTCA ACTAGCCGTT
      GcB5  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGAAT GCTAGTTGTC
```

```
Consensus  GGATTAGATA CCCTGGTAGT CCACGCCGTA AACGATGa.t gCTag.tGtt

           901                                                  950
     SdB3  TGGTGTCTTG AACAC.TGGC TTTCGGAGCT AACGCGTTAA GTAGACCGCC
     SdB4  CGGAACCTCG .GTTC.TGTT TTTCAAAGCT AACGCATTAA GTAGACCGCC
     GcB3  CGGAACCTCG .GTTC.TGTT TTTCAAAGCT AACGCATTAA GTAGACCGCC
    SdB11  CGGAACCTCG .GTTC.TGTT TTTCAAAGCT AACGCATTAA GTAGACCGCC
    SdB12  TGGTGCCTTG AGCGC.TGGG TGGCGCAGCT AACGCATTAA GTTGACCGCC
     SdB5  TGTGGCCTTG AGCCG.TGGC TTTCGGAGCT AACGCGTTAA GTAGACCGCC
    GcB13  TGTGGCCTTG AGCCG.TGGC TTTCGGAGCT AACGCGTTAA GTAGACCGCC
    GcB12  TGGTGCCTTG AGCAC.TGGG CTCCCAAGCT AACGCATTAA GTAGACCGCC
     GcB4  CGGAACCTCG .GTTC.TGTT TTTCAAAGCT AACGCATTAA GTAGACCGCC
    SdB10  AGGTAGCTTG CT.AT.TTGG TGACGCAGCT AACGCATTAA GCATTCCGCC
    GcB10  AGGTAGCTTG CT.AT.TTGG TGACGCAGCT AACGCATTAA GCATTCCGCC
     GcB9  GGGTCTCTTG CAGAC.TTGG TGACGAAGCT AACGCGATAA GTAGACCGCC
     GcB8  AGGTAGCTTG CTATT.TGG. TGACGCAGCT AACGCATTAA GCATTCCGCC
    GcB11  AGGTAGCTTG CTATT.TGG. TGACGCAGCT AACGCATTAA GCATTCCGCC
     SdB7  AGGGGGTTTC CGCCCCTTAG TGCTGCAGCT AACGCATTAA GCACTCCGCC
    SdB15  AGAGGGTTTC CGCCCTTTAG TGCTGCAGCT AACGCATTAA GCACTCCGCC
     SdB9  GGGGGGTT.C CACCC.TCAG TGCTGAAGTT AACACATTAA GCACTCCGCC
    SdB13  GGGGGGTT.C CACCC.TCAG TGCTGAAGTT AACACATTAA GCACTCCGCC
    SdB14  GGGGGGTT.C CACCC.TCAG TGCTGACGTT AACACATTAA GCACTCCGCC
     GcB6  GGGGGGTT.C CACCC.TCAG TGCTGAAGTT AACACATTAA GCACTCCGCC
     SdB8  AGAGGGTTTC CGCCCTTTAG TGCTGCAGCA AACGCATTAA GCACTCCGCC
     GcB7  GGGGGGTT.C CACCC.TCAG TGCTGAAGTT AACACATTAA GCACTCCGCC
     SdB6  GGGGATCTTG AATCC.TTAG TGGCGCAGCT AACGCACTAA GTTGACCGCC
     GcB5  AGGTAGCTTG CTAT..TTGG TGACGCAGCT AACGCATTAA GCATTCCGCC
Consensus  .Gg..gcTtg c...c.T.gg Tg.cg.AGcT AACgCatTAA Gca.tCCGCC

           951                                                 1000
     SdB3  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AATGAATTGA CGGGGGCCCG
     SdB4  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
     GcB3  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
    SdB11  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
    SdB12  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
     SdB5  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AATGAATTGA CGGGGGCCCG
    GcB13  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AATGAATTGA CGGGGGCCCG
    GcB12  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
     GcB4  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
    SdB10  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AAGGAATTGA CGGGGGCCCG
    GcB10  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AAG.AATTGA CGGGGGCCCG
     GcB9  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
     GcB8  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AAGGAATTGA CGGGGGCCCG
    GcB11  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AAGGAATTGA CGGGGGCCCG
     SdB7  TGGGGAGTAC GGTCGCAAGA CTGAAACTCA AAGGAATTGA CGGGGGCCCG
    SdB15  TGGGGAGTAC GGCCGCAAGG CTGAAACTCA AAGGAATTGA CGGGGGCCCG
     SdB9  TGGGGAGTAC GACCGCAAGG TTGAAACTCA AAGGAATTGA CGGGGGCCCG
    SdB13  TGGGGAGTAC GACCGCAAGG TTGAAACTCA AAGGAATTGA CGGGGGCCCG
    SdB14  TGGGGAGTAC GGCCGCAAGG CTGAAACTCA AAGGAATTGA CGGGGGCCCG
     GcB6  TGGGGAGTAC GACCGCAAGG TTGAAACTCA AAGGAATTGA CGGGGGCCCG
     SdB8  TGGGGAGTAC GACCGCAAGG TTGAAACTCA AAGGAATTGA CGGGGGCCCG
     GcB7  TGGGGAGTAC GACCGCAAGG TTGAAACTCA AAGGAATTGA CGGGGGCCCG
     SdB6  TGGGGAGTAC GGCCGCAAGG TTAAAACTCA AATGAATTGA CGGGGGCCCG
     GcB5  TGGGGAGTAC GGTCGCAAGA TTAAAACTCA AAGGAATTGA CGGGGGCCCG
Consensus  TGGGGAGTAC GgcCGCAAGg TTaAAACTCA AAgGAATTGA CGGGGGCCCG

           1001                                                1050
     SdB3  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     SdB4  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
```

```
      GcB3  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     SdB11  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     SdB12  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
      SdB5  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     GcB13  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     GcB12  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
      GcB4  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
     SdB10  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCTACGCG CAGAACCTTA
     GcB10  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCTACGCG CAGAACCTTA
      GcB9  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      GcB8  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG CAGAACCTTA
     GcB11  CACAAGCGGT GGAGCATGTG GTTTAATTCG ATGCAACGCG AAGAACCTTA
      SdB7  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
     SdB15  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      SdB9  CACAAGCAGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
     SdB13  CACAAGCAGT GGAGCATGTG GTTTAATTCG AAGCGACGCG AAGAACCTTA
     SdB14  CACAAGCAGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      GcB6  CACAAGCAGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      SdB8  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      GcB7  CACAAGCAGT GGAGCATGTG GTTTAATTCG AAGCAACGCG AAGAACCTTA
      SdB6  CACAAGCGGT GGAGCATGTG GTTTAATTCG ACGCAACGCG AAGAACCTTA
      GcB5  CACAAGCGGT GGAGCATGTG GTTTAATTCG AAGCAACGCG CAGAACCTTA
 Consensus  CACAAGCgGT GGAGCATGTG GTTTAATTCG A.GCAACGCG aAGAACCTTA

            1051                                                1100
      SdB3  CCTACTCTTG ACATCCAGAG AA.TTCGCTA GAGATAGCTC AGTGCC.TTC
      SdB4  CCTACACTTG ACATACAGAG AA.CTTACCA GAGATGGTTT GGTGCC.TTC
      GcB3  CCTACACTTG ACATACAGAG AA.CTTACCA GAGATGGTTT GGTGCC.TTC
     SdB11  CCTACACTTG ACATACAGAG AA.CTTACCA GAGATGGTTT GGTGCC.TTC
     SdB12  CCTACTCTTG ACATCCACAG AA.CTTTTCA GAGATGAATT GGTGCC.TTC
      SdB5  CCTACTCTTG ACATCCAGAG AA.CTTAGCA GAGATCGTTT GGTGCC.TTC
     GcB13  CCTACTCTTG ACATCCAGAG AA.GCCAGCG GAGACGCAGG TGTGCC.TTC
     GcB12  CCTACTCTTG ACATCCAGAG AA.TTCGCTA GAGATAGCTT AGTGCC.TTC
      GcB4  CCTACACTTG ACATACAGAG AA.CTTACCA GAGATGGTTT GGTGCC.TTC
     SdB10  CCAAGCCTTG ACATCCTTGG AATCTCGCAG AAACGCGAGA G.TGCC.TTC
     GcB10  CCAAGCCTTG ACATCCTTGG AATCTCGCAG AAACGCGAGA GGTGCC.TTC
      GcB9  CCTGGCCTTG ACATCCTGCG AA.CTTTCTA GAGATAGATT GGTGCC.TTC
      GcB8  CCAGCCCTTG ACATTTGACG CT.ACAACCG GAGACGGTTG GTTCCC.TTC
     GcB11  CCTGCCCTTG ACATACTGAG AA.CTTACCA GAGATGGTTT GGTGCC.TTC
      SdB7  CCAGGTCTTG ACATCCTCTG AC.AACCCTA GAGATAGGGC TTTCCC.TTC
     SdB15  CCAGGTCTTG ACATCCTCTG AC.AACCCTA GAGATAGGGC TTTCCCCTTC
      SdB9  CCAGGTCTTG ACATCCTCTG AC.AATCCTG GAGACAGGAC GTTCCCCTTC
     SdB13  CCAGGTCTTG ACATCCTCTG AC.AATCCTG GAGACAGGAC GTTCCCCTTC
     SdB14  CCAGGTCTTG ACATCCTTCG CT.ACTTCTA GAGATAGAAG GTTCCCCTTC
      GcB6  CCAGGTCTTG ACATCCTCTG AC.CACTCTA GAGATAGAGC TTTCCCCTTC
      SdB8  CCAGGTCTTG ACATCCTCTG AC.AATCCTA GAGATAGGAC TTTCCCCTTC
      GcB7  CCAGGTCTTG ACATCCTCTG AC.AATCCTG GAGACAGGAC GTTCCCCTTC
      SdB6  CCTGGCCTTG ACATGCAGAG AA.CTTTCCA GAGATGGATT GGTGCC.TTC
      GcB5  CCAGCCCTTG ACATTTGACG CT.ACAACCG GAGACGG.TT GGTTCCTTTC
 Consensus  CCtgc.CTTG ACATcc.gaG aa..tt.Cca GAGAt.G.tt ggTgCC.TTC

            1101                                                1150
      SdB3  GGGA.ACTCT GAGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
      SdB4  GGGA.ACTCT GATAC.GGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
      GcB3  GGGA.ACTCT GATACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
     SdB11  GGGA.ACTCT GATACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
     SdB12  GGGA.ACTGT GAGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
      SdB5  GGGA.ATTCT GAGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
     GcB13  GGGA.GCTCT GAGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
```

```
      GcB12  GGGA.ACTCT GAGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTTGTGA
       GcB4  GGGA.ACTCT GATACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTGCTGA
      SdB10  GGGA.ACCAG GTGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
      GcB10  GGGA.ACCAG GTGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
       GcB9  GGGA.ACGCA GTGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
       GcB8  GGGG.ACGTC AGGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
      GcB11  GGGA.ACTCA GATACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
       SdB7  GGGG.ACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
      SdB15  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
       SdB9  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCC CGTGTCGTGA
      SdB13  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
      SdB14  GGGGGACGAA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
       GcB6  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
       SdB8  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCACCT CGTGTCGTGA
       GcB7  GGGGGACAGA GTGACAGGTG GTGCATGGTT GTCGTCAGCT CGTGTCGTGA
       SdB6  GGGA.ACTCT GACACAGGTG CTGCATGGCC GTCGTCAGCT CGTGTCGTGA
       GcB5  GGGG.ACGTC AGGACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA
  Consensus  GGGa.ACtc. gagACAGGTG cTGCATGGcT GTCGTCAGCT CGTGTcGTGA

             1151                                                1200
       SdB3  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTAGTTGCTA
       SdB4  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTAGTTGCTA
       GcB3  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTAGTTGCTA
      SdB11  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTAGTTGCTA
      SdB12  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTAGTTGCCA
       SdB5  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTATCC TTGTTTGCCA
      GcB13  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTATCC TTGTTTGCCA
      GcB12  AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTATTTGCCA
       GcB4  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTAAGC TTAGTTGCCA
      SdB10  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC CTATTTGCCA
      GcB10  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC CTATTTGCCA
       GcB9  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGTCC TTAGTTACCA
       GcB8  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTCGCCC TTAGTTGCCA
      GcB11  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCCTATCC TTATTTGCCA
       SdB7  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
      SdB15  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
       SdB9  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
      SdB13  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
      SdB14  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
       GcB6  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGACC TTAGTTGCCA
       SdB8  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
       GcB7  GATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTGATC TTAGTTGCCA
       SdB6  GATGTTGGGT TAAGTCCCGT AACGAGCGCA ACCCCTATCC TTGGTTGCTA
       GcB5  GATGT.GGGT TAAGTCCCGC AACGAGCGCA ACCCTCGCCC TTAGTTGCCA
  Consensus  gATGTtGGGT TAAGTCCCGC AACGAGCGCA ACCCttatcC TTagTTGCcA

             1201                                                1250
       SdB3  GCAGGTAATG CTGAGAACTC TAAGGAGACT GCCGGTGATA AACCGGAGGA
       SdB4  GCAGGTAATG CTGAGAACTC TAAGGAGACT GCCGGTGATA AACCGGAGGA
       GcB3  GCAGGTAATG CTGAGAACTC TAAGGAGACT GCCGGTGATA AACCGGAGGA
      SdB11  GCAGGTAATG CTGAGAACTC TAAGGAGACT GCCGGTGATA AACCGGAGGA
      SdB12  GCGAGTAATG TCGGGAACTC TAGGGAGACT GCCGGTGATA AACCGGAGGA
       SdB5  GCGAGTAATG TCGGGAACTC CAGGGAGACT GCCGGTGATA AACCGGAGGA
      GcB13  GCGAGTCATG TCGGGAACTC CAGGGAGACT GCCGGTGATA AACCGGAGGA
      GcB12  GCACGTTATG GTGGGAACTC TAGGGAGACT GCCGGTGATA AACCGGAGGA
       GcB4  TCA..TTAAG TTGGGCACTC TAAGTTGACT GCCGGTGACA AACCGGAGGA
      SdB10  GCAC.TTCGG GTGGGAACTT TAGGGAGACT GTCGGTGATA ATTCGAAGGA
      GcB10  GCAC.TTCGG GTGGGAACTT TAGGGAGACT GTCGGTGATA AT.CGAAGGA
       GcB9  GCACGTTATG GTGGGCACTC TAGGGAGACT GCCGGTGACA AACCGGAGGA
```

```
      GcB8  GCATTTAGT. .TGGGCACTC TAGGGGGACT GCCGGTGATA AGCCGGAGGA
     GcB11  GCACGTAATG GTGGGAACTC TAGGGAGACT GCCGGTGATA AACCGGAGGA
      SdB7  GCA..TTTAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
     SdB15  GCA..TTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
      SdB9  GCA..TTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
     SdB13  GCA..TTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
     SdB14  GCA..TTTAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
      GcB6  GCA..TTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
      SdB8  GCA..TTTAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
      GcB7  GCA..TTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA
      SdB6  GCAGGTAATG CTGAGAACTC CAGGGAGACT GCCGGTGACA AACCGGAGGA
      GcB5  GCA..TTTAG TTGG...... .......... .......... ..........
Consensus  GCA..TtatG .TGggaactc ta.ggagact gccggtga.a aaccggagga

           1251                                                1300
      SdB3  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGTGTA GGGCTACACA
      SdB4  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGTGTA GGGCTACACA
      GcB3  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGTGTA GGGCTACACA
     SdB11  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGTGTA GGGCTACACA
     SdB12  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGTGTA GGGCTACACA
      SdB5  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGAGTA GGGCTACACA
     GcB13  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGAGTA GGGCTACACA
     GcB12  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGAGTA GGGCTACACA
      GcB4  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGATTT GGGCTACACA
     SdB10  AGGTGGGGAT GACGTCAAGT CATCATGGCC CTTATGGCTT GGGCTACACA
     GcB10  AGGTGGGGAT GACGTCAAGT CATCATGGCC CTTATGGCTT GGGCTACACA
      GcB9  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGGCCA GGGCTACACA
      GcB8  AGGTGGGGAT GACGTCAAGT CCTCATGGCC CTTACGGGCT GGGCTACACA
     GcB11  AGGTGGGGAC GACGTCAAGT CATCATGGCC CTTACGGGCA GGGCTACACA
      SdB7  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
     SdB15  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
      SdB9  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
     SdB13  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
     SdB14  AGGTGGGGAC GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
      GcB6  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
      SdB8  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
      GcB7  AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA
      SdB6  AGGTGGGGAT GACGTCAGGT CATCATGGCC CTTACGGCCA GGGCTACACA
      GcB5  .......... .......... .......... .......... ..........
Consensus  aggtggggga. gacgtcaagt catcatggcc cttacg.... gggctacaca

           1301                                                1350
      SdB3  C.TGCTACAA TGGCGCATAC AGAGTGCTGC GAACCTGCGA AGGTAAGCGA
      SdB4  CGTGCTACAA TGGCGCATAC AGAGTGCTGC GAACCTGCGA AGGTAAGCGA
      GcB3  CGTGCTACAA TGGCGCATAC AGAGTGCTGC GAACTCGCGA GAGTAAGCGA
     SdB11  CGTGCTACAA TGGCGCATAC AGAGTGCTGC GAACTCGCGA GAGTAAGCGA
     SdB12  CGTGCTACAA TGGCGCATAC AGAGTGCTGC GAACTCGCGA GAGTAAGCGA
      SdB5  CGTGCTACAA TGGCGCATAC AGAGGGCAGC CAACTTGCGA AAGTGAGCGA
     GcB13  CGTGCTACAA TGGCGCATAC AGAGGGCAGC AAGCTAGCGA TAGTGAGCGA
     GcB12  CGTGCTACAA TGGCGTATAC AGAGGGTTGC AAAGCCGCAA GGTCTAGCTA
      GcB4  CGTGCTACAA TGGACAATAC AAAGGGCAGC TAAACCGCGA GGCCAAGCAA
     SdB10  CGTGCTACAA TGGCCGGTAC AATAGGTCGC TAACCCGCGA GGGGGAGTAA
     GcB10  CGTGCTACAA TGGCCGGTAC AATAGGTCGC TAACCCGCGA GGGGGAGTAA
      GcB9  CGTGCTACAA TGGTGCATAC AGACGGTTGC CAAGCCGCGA GGTGGAGCTA
      GcB8  CGTGCTACAA TGGCGGTGAC AGTGGGCAGC GACCTCGCGA GGGGAAGCTA
     GcB11  CGTGCTACAA TGGCATGTAC AGAGGGATGC GAACTCGCGA GAGCAAGCGG
      SdB7  CGTGCTACAA TGGACAGAAC AAAGGGCTGC GAGACCGCAA GGTTTAGCCA
     SdB15  CGTGCTACAA TGGATGGTAC AAAGGGCTGC AAACCTGCGA AGGTAAGCGA
      SdB9  CGTGCTACAA TGGACGGTAC AAAGGGCAGC AACACCGCGA GGTGAAGCGA
```

```
      SdB13  CGTGCTACAA TGGACGGTAC AAAGGGCAGC AACACCGCGA GGTGAAGCGA
      SdB14  CGTGCTACAA TGGATGGTAC AAAGGGCAGC AAAACCGCGA GGTTGAGCGA
       GcB6  CGTGCTACAA TGGATGGTAC AAAGGGTTGC GAAGCCGCGA GGCCAAGCCA
       SdB8  CGTGCTACAA TGGATGGTAC AAAGGGCTGC AAGACCGCGA GGTTTAGCCA
       GcB7  CGTGCTACAA TGGACGGTAC AAAGGGCAGC AACACCGCGA G.........
       SdB6  CGTGCTACAA TGGCGTATAC AGAGGGCTGC CAACTCGCGA GAGTGAGCCA
       GcB5  .......... .......... .......... .......... ..........
  Consensus  cgtgctacaa tggc...tac agagggctgc .aacccgcga gggt.agc.a

             1351                                                1400
      SdB3   ATCACTTAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
      SdB4   ATCACTTAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
       GcB3  ATCACTTAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
      SdB11  ATCACTTAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
      SdB12  ATCACTTAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CCCGACTCCA
       SdB5  ATCCCAAAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
      GcB13  ATCCCAAAAA GTGCGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
      GcB12  ATCTCACAAA GTACGTCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
       GcB4  ATCCCATAAA GTTGTTCTCA GTTCGGATTG TAGTCTGCAA CTCGACTACA
      SdB10  ATCCGAAAAA GCCGGTCGTA GTCCGGATCG AAGTCTGCAA CTCGACTTCG
      GcB10  ATCCGAAAAA GCCGGTCGTA GTCCGGATCG AAGTCTGCAA CTCGACTTCG
       GcB9  ATCTGAGAAA GTGCATCGTA GTCCGGATTG GAGTCTGCAA CTCGACTCCA
       GcB8  ATCTCTAAAA GCC.GTCTCA GTTCGGATTG TTCTCTGCAA CTCGAGAGCA
      GcB11  ACCCCAAAAA GCATGTCGTA GTCCGGATCG GAGTCTGCAA CTCGACTCCG
       SdB7  ATCCCACAAA TCTGTTCTCA GTTCGGATCG CAGTCTGCAA CTCGACTGCG
      SdB15  ATCCCATAAA GCCATTCTCA GTTCGGATTG CAGGCTGCAA CTCGCCTGCA
       SdB9  ATCCCATAAA GCCGTTCTCA GTTCGGATTG CAGGCTGCAA CTCGCCTGCA
      SdB13  ATCCCATAAA GCCGTTCTCA GTTCGGATTG CAGGCTGCAA CTCGCCTGCA
      SdB14  ATCCCATAAA GCCATTCTCA GTTCGGATTG CAGGCTGCAA CTCGCCTGCA
       GcB6  ATCCCAAAAA GCCATTCTCA GTTCGGATTG TAGGCTGCAA CTCGCCTACA
       SdB8  ATCCCATAAA ACCATTCTCA GTTCGGATTG TAGGCTGCAA CTCGCCTACA
       GcB7  .......... .......... .......... .......... ..........
       SdB6  ATCCCTTAAA GTGCGTCGTA GTCCGGATCG CAGTCTGCAA CTCGACTGCG
       GcB5  .......... .......... .......... .......... ..........
  Consensus  atc.c.taaa g...gtcgta gtccggattg .agtctgcaa ctcgact.ca

             1401                                                1450
      SdB3   TGAAGTCGGA ATCGCTAGTA ATCGCGTATC AG.AATGACG CGGTGAATAC
      SdB4   TGAAGTCGGA ATCGCTAGTA ATCGCGTATC AG.AATGACG CGGTGAATAC
       GcB3  TGAAGTCAGA ATCGCTAGTA ATCGCGTATC AG.AATGACG CGGTGAATAC
      SdB11  TGAAGTCGGA ATCGCTAGTA ATCGCGTATC AG.AATGACG CGGTGAATAC
      SdB12  TGAAGTCGGA ATCGCTAGTA ATCGCATATC AG.AATGATG CGGTGAATAC
       SdB5  TGAAGTCGGA ATCGCTAGTA ATCGTGGATC AG.AATGCCA CGGTGAATAC
      GcB13  TGAAGTCGGA ATCGCTAGTA ATCGTGAATC AG.AATGTCA CGGTGAATAC
      GcB12  TGAAGTCGGA ATCGCTAGTA ATCGTAGATC AG.AATGCTA CGGTGAATAC
       GcB4  TGAAGCTGGA ATCGCTAGTA ATCGTAGATC AG.CATGCTA CGGTGAATAC
      SdB10  TGAAGTCGGA ATCGCTAGTA ATCGCGAATC AGCAATGTCG CGGTGAATAC
      GcB10  TGAAGTCGGA ATCGCTAGTA ATCGCGAATC AGCAATGTCG CGGTGAATAC
       GcB9  TGAAGTCGGA ATCGCTAGTA ATCGTGAATC AG.AATGTCA CGGTGAATAC
       GcB8  TGAAGTTGGA ATCGCTAGTA ATCGCGTAAC AG.CATGACG CGGTGAATAC
      GcB11  TGAAGTCGGA ATCGCTAGTA ATCGTAGATC AG.AATGCTA CGGTGAATAC
       SdB7  TGAAGCTGGA ATCGCTAGTA ATCGCGGATG AG.CATGCCG CGGTGAATAC
      SdB15  TGAAGCCGGA ATCGCTAGTA ATCGCGGATC AG.CATGCCG CGGTGAATAC
       SdB9  TGAAGCCGGA ATTGCTAGTA ATCGCGGATC AG.CATGCCG CGGTGAATAC
      SdB13  TGAAGCCGGA ATTGCTAGTA ATCGCGGATC AG.CATGCCG CGGTGAATAC
      SdB14  TGAAGCCGGA ATTGCTAGTA ATCGCGGATC AG.CATGCCG CGGTGAATAC
       GcB6  TGAAGCCGGA ATTGCTAGTA ATCGCGGATC AG.CATGCCG CGGTGAATAC
       SdB8  TGAAG..... .......... .......... .......... ..........
       GcB7  .......... .......... .......... .......... ..........
```

```
      SdB6  TGAAGTCGGA ATCGCTAGTA ATCGCGAATC AG.AATGTCG CGGTGAATAC
      GcB5  .......... .......... .......... .......... ..........
Consensus  tgaagtcgga atcgctagta atcgcg.atc ag.aatg.cg cggtgaatac

           1451                                                1500
      SdB3  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
      SdB4  GTTCCCGGGC CTTGCACACA CCG..GCCGA TTCCAGCACA CTGGCGGCCG
      GcB3  GTTCCCGGGC CTTGCACACA CC........ .......... ..........
     SdB11  G......... .......... .......... .......... ..........
     SdB12  GTTCCCGGGC CTTGCACACA CCGAAGCCGA A......... ..........
      SdB5  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
     GcB13  GTTCCCGGGC CTTGTACACA CCG....... .......... ..........
     GcB12  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
      GcB4  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
     SdB10  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
     GcB10  GTTCCCGGGC CTTGCACAC. .......... .......... ..........
      GcB9  GTTCCCGGGC CTTGCACACA CCG....... .......... ..........
      GcB8  GTTCCCGGGC CTTGCACACA CC........ .......... ..........
     GcB11  GTTCCCGGGC CTTGTACACA CCGAAGCC.. .......... ..........
      SdB7  GTTCCCGGGC CTTGCACACA CC........ .......... ..........
     SdB15  GTTCCCGGGC CTTGCACACA CCGAAGCC.. .......... ..........
      SdB9  GTTCCCGGGC CTTGCACACA CCGAAGC... .......... ..........
     SdB13  GTTCCCGGGC CTTGCACACA CC........ .......... ..........
     SdB14  GTTCCCGGGC CTTGTACACA CCG....... .......... ..........
      GcB6  GTTCCCGGGC CTTGTACACA CC........ .......... ..........
      SdB8  .......... .......... .......... .......... ..........
      GcB7  .......... .......... .......... .......... ..........
      SdB6  GTTCCCGGGC CTTGTACACA CCG....... .......... ..........
      GcB5  .......... .......... .......... .......... ..........
Consensus  gttcccgggc cttgcacaca ccg....... .......... ..........
```

Appendix B

Alignment of the 18 sequences of Archaea clones utilized for the phylogenetic three showed in Figs.3.33-3.34-3.35.

```
Multalin version 5.4.1
Copyright I.N.R.A. France 1989, 1991, 1994, 1996
Published research using this software should cite
Multiple sequence alignment with hierarchical clustering
F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890
Symbol comparison table: blosum62
Gap weight: 12
Gap length weight: 2
Consensus levels: high=90% low=50%
Consensus symbols:
 ! is anyone of IV
 $ is anyone of LM
 % is anyone of FY
 # is anyone of NDQEBZ

 MSF:   1453    Check:   0          ..
 Name: AJ347776       Len:  1453  Check: 6697  Weight:  0.55
 Name: AJ347774       Len:  1453  Check: 7786  Weight:  0.55
 Name: GcAr1          Len:  1453  Check:  649  Weight:  0.88
 Name: GcAr11         Len:  1453  Check:  961  Weight:  0.88
 Name: GcAr13         Len:  1453  Check:  604  Weight:  0.88
 Name: GcAr18         Len:  1453  Check: 1029  Weight:  0.88
 Name: GcAr2          Len:  1453  Check: 1862  Weight:  0.91
 Name: GcAr17         Len:  1453  Check:  389  Weight:  0.91
 Name: GcAr4          Len:  1453  Check:  741  Weight:  0.91
 Name: GcAr3          Len:  1453  Check: 1352  Weight:  0.91
 Name: GcAr8          Len:  1453  Check:  672  Weight:  0.91
 Name: GcAr14         Len:  1453  Check: 2144  Weight:  0.91
 Name: GcAr15         Len:  1453  Check: 5042  Weight:  0.91
 Name: GcAr7          Len:  1453  Check: 2336  Weight:  0.91
 Name: GcAr12         Len:  1453  Check: 4589  Weight:  0.88
 Name: GcAr16         Len:  1453  Check: 3608  Weight:  0.88
 Name: GcAr9          Len:  1453  Check: 2097  Weight:  0.91
 Name: GcAr10         Len:  1453  Check: 3775  Weight:  0.94
 Name: GcAr5          Len:  1453  Check: 9895  Weight:  1.37
 Name: GcAr6          Len:  1453  Check: 3881  Weight:  3.07
 Name: Consensus      Len:  1453  Check: 1029  Weight:  0.00

//

                    1                                                  50
          AJ347776  .TCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
          AJ347774  ..CTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr1  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr11  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr13  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr18  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr2  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr17  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr4  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr3  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr8  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr14  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr15  TXXXGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr7  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr12  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr16  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
             GcAr9  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
            GcAr10  TXXXGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
```

```
    GcAr5  TTCCGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
    GcAr6  TTCTGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA
Consensus  TtCtGGTTGA TCCTGCCGGA CCTGACTGCT ATCGGATTGA TACTAAGCCA

           51                                                  100
 AJ347776  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
 AJ347774  TGCGAGTCAT TGTAGCAATA CAAGGCATAC GGCTCAGTAA CGCGTAGTCA
    GcAr1  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr11  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr13  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr18  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr2  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr17  TGCGAGCCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr4  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCATAGTCA
    GcAr3  TGCGAGTCAT TGTAGTAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr8  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr14  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr15  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr7  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr12  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr16  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr9  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
   GcAr10  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr5  TGCGAGTCAT TGTAGTAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
    GcAr6  TGCGAGTCAT TGTAGCAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA
Consensus  TGCGAGTCAT TGTAGcAATA CAAGGCAGAC GGCTCAGTAA CGCGTAGTCA

           101                                                 150
 AJ347776  ACCTAACCTA TGGACGGGAA TAACCTCGGG AAACTGAGAA TAATGCCCGA
 AJ347774  ACCTAACCTA TGGACGGGAA TAACCTCGGG AAACTGAGAA TAATGCCCGA
    GcAr1  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr11  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr13  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr18  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr2  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr17  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr4  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr3  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr8  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr14  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr15  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr7  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr12  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
   GcAr16  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr9  ACCTACCCTA TGGACGGGAA TAACCTCGGG AGACTGAGTA TAATGCCCGA
   GcAr10  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr5  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
    GcAr6  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA
Consensus  ACCTACCCTA TGGACGGGAA TAACCTCGGG AAACTGAGTA TAATGCCCGA

           151                                                 200
 AJ347776  TAGAACACTA TACCTGGAAT GGTTTGTGTT CCAAATGATT TATCGCCGTA
 AJ347774  TAGAACATTG TGCCTGGAAT GGTTTATGTT CCAAATGATT TATCGCCGTA
    GcAr1  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
   GcAr11  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
   GcAr13  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
   GcAr18  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
    GcAr2  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
   GcAr17  TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
```

```
GcAr4      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr3      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr8      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr14     TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr15     TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr7      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr12     TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr16     TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr9      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr10     TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
GcAr5      TAGAACACAA TGCCTGGAAC GGTTTGTGTT CCAAATGATT TATCGCCGTA
GcAr6      TAGAACACTA TGCCTGGAAA GGTTTGTGTT CCAAACGATT TATCGCCGTA
Consensus  TAGAACACTA TGCCTGGAAa GGTTTGTGTT CCAAAcGATT TATCGCCGTA

           201                                                 250
AJ347776   GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
AJ347774   GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr1      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr11     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr13     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr18     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr2      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr17     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr4      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr3      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr8      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr14     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr15     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr7      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr12     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr16     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr9      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr10     GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr5      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
GcAr6      GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG
Consensus  GGATGGGACT GCGGCCTATC AGTTTGTTGG TGAGGTAATG GCCCACCAAG

           251                                                 300
AJ347776   ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
AJ347774   ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr1      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr11     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr13     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr18     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr2      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr17     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr4      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr3      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr8      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr14     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr15     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr7      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr12     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr16     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr9      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr10     ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr5      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
GcAr6      ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
Consensus  ACTATTACAG GTACGGGCTC TGAGAGGAGT AGCCCGGAGA TGGGTACTGA
```

```
           301                                                       350
AJ347776   GACACGGACC CAGGACCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
AJ347774   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr1   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr11   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr13   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr18   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr2   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr17   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr4   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr3   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr8   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr14   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr15   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr7   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr12   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr16   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr9   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
  GcAr10   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr5   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
   GcAr6   GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT
Consensus  GACACGGACC CAGGCCCTAT GGGGCGCAGC AGGCGAGAAA ACTTTGCAAT

           351                                                       400
AJ347776   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
AJ347774   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr1   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr11   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr13   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr18   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr2   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr17   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr4   GTGCGAAAGC ACGGCAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr3   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr8   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr14   GTGCGAAAGC ACGACAAGGT AAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr15   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr7   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr12   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr16   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr9   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
  GcAr10   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr5   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
   GcAr6   GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT
Consensus  GTGCGAAAGC ACGACAAGGT TAATCCGAGT GGTTTCTGCT AAAGGAACCT

           401                                                       450
AJ347776   TTTGTTAGTC CTAGAAACAC TAACGAATAA GGGGTGGGCA AGTTCTGGTG
AJ347774   TTTGTCAGTC CTAGAAACAC TGATGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr1   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
  GcAr11   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
  GcAr13   TTTGACAGTC CTAAAAACAC TGTCGAATAA GGGGTGGGCA AGTTCTGGTG
  GcAr18   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr2   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
  GcAr17   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr4   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr3   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr8   TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
```

```
   GcAr14  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr15  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
    GcAr7  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr12  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr16  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
    GcAr9  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
   GcAr10  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
    GcAr5  TTTGACAGTC CTAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
    GcAr6  TTTGACAGTC CCAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG
Consensus  TTTGACAGTC CtAAAAACAC TGTTGAATAA GGGGTGGGCA AGTTCTGGTG

           451                                                 500
 AJ347776  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
 AJ347774  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr1  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr11  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr13  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr18  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr2  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr17  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr4  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr3  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAC.TAT.
    GcAr8  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr14  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGATATATA
   GcAr15  TTATCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGATATATA
    GcAr7  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr12  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGATATATA
   GcAr16  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr9  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
   GcAr10  TCATCCGCCG CGGTAAAAAC CAGCACCTCA AGTGGTCAGG ATGATATATA
    GcAr5  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
    GcAr6  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.
Consensus  TCAGCCGCCG CGGTAAAA.C CAGCACCTCA AGTGGTCAGG ATGAT.TAT.

           501                                                 550
 AJ347776  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT AAGTTTTCGG TTAAATCTGT
 AJ347774  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT AAGTTTTCGG TTAAATCTGT
    GcAr1  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr11  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr13  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr18  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr2  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr17  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr4  TGGGCCTAAA ACATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr3  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr8  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr14  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr15  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr7  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr12  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr16  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr9  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
   GcAr10  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr5  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
    GcAr6  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT
Consensus  TGGGCCTAAA GCATCCGTAG CCGGCTCTGT TAGTTTTCGG TTAAATCTGT

           551                                                 600
 AJ347776  ACGCTCAACG TACAGGCTGC CGGGAATACT GCAAAGCTAG GGAGTGGGAG
```

```
AJ347774  ACGCTCAACG TACAGGCTGC CGGGAATACT GCATAGCTAG GGAGTGGGAG
   GcAr1  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr11  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr13  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr18  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr2  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr17  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr4  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr3  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr8  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr14  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr15  ACGCTCACCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr7  ACGCTCACCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr12  ACGCTCACCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr16  ACGCTCACCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr9  ACGCTCACCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
  GcAr10  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
   GcAr5  GCGCTCAACG TACAGGCTGC CGGAAATACT ACAGAGCTAG GGAGTGGGAG
   GcAr6  ACGCTCAACG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG
Consensus ACGCTCAaCG TACAGGCTGC CGGAAATACT GCAGAGCTAG GGAGTGGGAG

          601                                                 650
AJ347776  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
AJ347774  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr1  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr11  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr13  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr18  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr2  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr17  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr4  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr3  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr8  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr14  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
  GcAr15  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr7  AGGTAGACGG TACTCGGTAG GAAGGGGTGA GAGATATCCT TTGATCTATT
  GcAr12  AGGTAGACGG TACTCGGTAG GAAGGGGTGA GAGATATCCT TTGATCTATT
  GcAr16  AGGTAGACGG TACTCGGTAG GAAGGGGTGA GAGATATCCT TTGATCTATT
   GcAr9  AGGTAGACGG TACTCGGTAG GAAGGGGTGA GAGATATCCT TTGATCTATT
  GcAr10  AGGTAGACGG TACTCGGTAG GAAGGGGTGA GAGATATCCT TTGATCTATT
   GcAr5  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
   GcAr6  AGGTAGACGG TACTCGGTAG GAAGGGGTAA AA....TCCT TTGATCTATT
Consensus AGGTAGACGG TACTCGGTAG GAAGGGGTaA aA....TCCT TTGATCTATT

          651                                                 700
AJ347776  GATGACC.CC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
AJ347774  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr1  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
  GcAr11  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
  GcAr13  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
  GcAr18  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr2  GATGACCACC TGTGGCGAAG GCGGTTTACC AGAACACGTC CG.ACGGTGA
  GcAr17  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr4  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr3  GATGACCACC TGTGGCGAAG GCGGTTTACC AGAACACGTC CG.ACGGTGA
   GcAr8  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
  GcAr14  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
  GcAr15  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr7  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
```

```
   GcAr12  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
   GcAr16  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
    GcAr9  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CGTACGGTGA
   GcAr10  GATGACCAC. TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
    GcAr5  GATGACCACC TGTGGCGAAG GCGGTTTACC AGAACACGTC CG.ACGGTGA
    GcAr6  GATGACCACC TGTGGCGAAG GCGGTCTACC AGAACACGTC CG.ACGGTGA
Consensus  GATGACCACC TGTGGCGAAG GCGGTcTACC AGAACACGTC CG.ACGGTGA

           701                                                  750
 AJ347776  GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
 AJ347774  GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr1   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr11   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr13   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr18   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr2   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr17   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr4   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr3   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr8   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr14   GGGATGGAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr15   GGGATGAAAG CTGGGGGAGC AAACCGGATT ATATACCCGG GTAGTCCCAG
   GcAr7   GGGATGAAAG CTGGGGGAGC AAACCGGATT ATATACCCGG GTAGTCCCAG
  GcAr12   GGGATGAAAG CTGGGGGAGC AAACCGGATT ATATACCCGG GTAGTCCCAG
  GcAr16   GGGATGAAAG CTGGGGGAGC AAACCGGATT ATATACCCGG GTAGTCCCAG
   GcAr9   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
  GcAr10   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr5   GGGATGAAAG CTGGGGGAGC AAACCGGATT AGATACCCGG GTAGTCCCAG
   GcAr6   GGGATGAA.. .......... .......... .......... ..........
Consensus  GGGATGAAag ctgggggagc aaaccggatt agatacccgg gtagtcccag

           751                                                  800
 AJ347776  CTGTAAACTA TGC.AAACTC AGTGATGCAT TGGGCTTGTG GCCAATGCAG
 AJ347774  CTGTAA.CTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG .CCAATGCAG
   GcAr1   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr11   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr13   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr18   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr2   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr17   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr4   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr3   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr8   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr14   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr15   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr7   CTGTAAACTA TGCATAACTC AGTGATGCAT TGG.CTTGTG GCCA.TGCAG
  GcAr12   CTGTAAACTA TGC.TAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr16   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr9   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
  GcAr10   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr5   CTGTAAACTA TGC.AAACTC AGTGATGCAT TGG.CTTGTG GCCAATGCAG
   GcAr6   .......... .......... .......... .......... ..........
Consensus  ctgtaaacta tgc.aaactc agtgatgcat tgg.cttgtg gccaatgcag

           801                                                  850
 AJ347776  TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACCCAAGT
 AJ347774  TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
   GcAr1   TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
  GcAr11   TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
```

```
GcAr13    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr18    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr2     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr17    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr4     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr3     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr8     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr14    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr15    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr7     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr12    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr16    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr9     TGCTGCAGGG A.GCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr10    TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr5     TGCTGCAGGG AAGCCGTTAA GTTTGCCGCC TGGGAAGTAC GTACGCAAGT
GcAr6     .......... .......... .......... .......... ..........
Consensus tgctgcaggg aagccgttaa gtttgccgcc tgggaagtac gtacgcaagt

          851                                                  900
AJ347776  ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
AJ347774  ATGAAACTTA AAGGAATTGG CGGGG.AGCA CCACAAGGGG TGAAGCCTGC
GcAr1     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr11    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr13    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr18    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr2     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr17    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACATAGGG TGAAGCCTGC
GcAr4     ATGAAACTTA AAGGAATTGG CGGGGGGGCA CCACAAGGGG TGAAGCCTGC
GcAr3     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr8     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr14    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr15    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr7     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr12    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr16    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr9     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr10    ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr5     ATGAAACTTA AAGGAATTGG CGGGGGAGCA CCACAAGGGG TGAAGCCTGC
GcAr6     .......... .......... .......... .......... ..........
Consensus atgaaactta aaggaattgg cgggggagca ccacaagggg tgaagcctgc

          901                                                  950
AJ347776  GGTTC.AATT GGAGTCAACG CCAAAAATCT TACCCGGAGA GACAGCAGAA
AJ347774  GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr1     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr11    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr13    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr18    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr2     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr17    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr4     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr3     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr8     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGG GACAGCAGAA
GcAr14    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr15    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr7     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr12    GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr16    GGTTCCAATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
GcAr9     GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
```

```
   GcAr10  GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
    GcAr5  GGTTC.AATT GGAGTCAACG CCAGAAATCT TACCCGGAGA GACAGCAGAA
    GcAr6  .......... .......... .......... .......... ..........
Consensus  ggttc.aatt ggagtcaacg ccagaaatct tacccggaga gacagcagaa

           951                                                1000
 AJ347776  TGAAGGTCAA GCTGAAGACT TTACCAGACA AGCTGAGAGG TGGTGCATGG
 AJ347774  TGAAGGTCAA GCTGAAGACT TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr1  TGAAGGTCAG GCTAAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr11  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr13  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr18  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr2  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr17  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr4  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr3  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr8  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAG. TGGTGCATGG
   GcAr14  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr15  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr7  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr12  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr16  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr9  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
   GcAr10  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGCATGG
    GcAr5  TGAAGGTCAG GCTGAAGACC TTACCAGACA AGCTGAGAGG TGGTGC....
    GcAr6  .......... .......... .......... .......... ..........
Consensus  tgaaggtcag gctgaagacc ttaccagaca agctgagagg tggtgcatgg

           1001                                               1050
 AJ347776  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
 AJ347774  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCGG GTAACGAGCG
    GcAr1  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr11  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr13  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr18  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr2  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr17  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr4  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr3  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr8  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr14  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr15  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr7  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr12  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr16  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr9  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
   GcAr10  CCGTCGCCAG CTCGTGCCGT GAGATGTCCT GTTAAGTCAG GTAACGAGCG
    GcAr5  .......... .......... .......... .......... ..........
    GcAr6  .......... .......... .......... .......... ..........
Consensus  ccgtcgccag ctcgtgccgt gagatgtcct gttaagtcag gtaacgagcg

           1051                                               1100
 AJ347776  AGATCCCTGC CTCTAGTTGC CTCCATTACT CTCAGGAGTA GTGGGGCGAA
 AJ347774  AGATCCCTGC CTCTAGTTGC CACCATTACT CTCAGGAGTA GTGGGGCGAA
    GcAr1  AGA....... .......... .......... .......... ..........
   GcAr11  AGA....... .......... .......... .......... ..........
   GcAr13  AGA....... .......... .......... .......... ..........
   GcAr18  AGA....... .......... .......... .......... ..........
    GcAr2  AGA....... .......... .......... .......... ..........
```

```
GcAr17    AGA..... ..... ..... ..... .....
 GcAr4    AGA..... ..... ..... ..... .....
 GcAr3    AGA..... ..... ..... ..... .....
 GcAr8    AGA..... ..... ..... ..... .....
GcAr14    AGA..... ..... ..... ..... .....
GcAr15    AGA..... ..... ..... ..... .....
 GcAr7    AGA..... ..... ..... ..... .....
GcAr12    AGA..... ..... ..... ..... .....
GcAr16    AGA..... ..... ..... ..... .....
 GcAr9    AGA..... ..... ..... ..... .....
GcAr10    AGA..... ..... ..... ..... .....
 GcAr5    ..... ..... ..... ..... .....
 GcAr6    ..... ..... ..... ..... .....
Consensus aga..... ..... ..... ..... .....
```